UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE INFORMÁTICA

RoboScout: Automação de processo proativo de pesquisa, análise e tomada de decisão, com base em "Open-Source Intelligence"

Bruno Rafael Vilares Teixeira

Mestrado em Segurança Informática

Trabalho de Projeto orientado por:
Prof. Doutor Pedro Miguel Frazão Fernandes Ferreira

2021

Agradecimentos

Em primeiro lugar, queria agradecer aos meus pais, Teresa e António. Por todo o esforço que tiveram e sacrifícios que fizeram para hoje estar a escrever esta tese. Agradeço a ti mãe, por teres estado sempre ao meu lado e me teres tornado numa pessoa lutadora. Agradeço a ti pai, pelo sacrifício que tiveste para me dares uma educação e me teres ensinado a dar valor às pequenas coisas da vida. Nunca vou conseguir agradecer o suficiente por tudo o que fizeram por mim.

Também quero agradecer à minha namorada, Maria. Foi ela que esteve em todos os momentos, altos e baixos, e que me ajudou neste percurso. Sem ela, as coisas teriam sido muito mais difíceis. Obrigado por tudo o que fizeste por mim. Amo-te.

Quero agradecer a toda a equipa da DCY-CWM, especialmente ao meu principal apoio no projeto, Alfredo Alvim. Pelas palavras que me deu ao longo do ano, que me inspiraram a tornar-me numa melhor pessoa e profissional. Sem a sua ajuda e persistência nada disto seria possível.

Agradeço ao meu orientador Professor Pedro Ferreira, pela ajuda que me deu ao longo do ano. Mesmo com poucos contactos, foi essencial para me guiar na tese.

Agradeço ao meu coorientador, Engº José Alegria. Por ser incansável na procura do melhor e por me ter guiado pelo projeto. Foi um privilégio ser seu mestrando.

Agradeço à Sara Nascimento e ao Joel Jacinto por terem tido sempre tempo para me ajudar. O vosso apoio foi fulcral para o projeto. Obrigado ao Jorge Vigário por todo o apoio e por ter feito este caminho comigo.

E finalmente quero agradecer a todos os meus amigos. Aos que conheci na FCUL, um muito obrigado por estes anos e vivencias que levo para a vida. Aos da minha terra, obrigado por estarem sempre presentes.

A todos os que me ajudaram e acreditaram em mim

Resumo

Monitorizar a exposição de sistemas e serviços corporativos expostos na Internet é um processo fundamental para garantir a segurança e estabilidade da rede. A superfície de ataque, ou perímetro, a ser verificada é exponencial ao tamanho e heterogeneidade da estrutura empresarial, o que pode dificultar a verificação em empresas de grande escala, ou que possuam um grande número de serviços.

Para além dos sistemas e serviços, a superfície também engloba a deteção de informação exposta na Internet que seja considerada crítica para a empresa. A utilização de fontes OSINT que coletem diferentes tipos de informação acerca do perímetro empresarial, seja dos serviços ou de informação, pode ser organizada numa forma estruturada e normalizada que permita uma aferição da proveniência desses dados. Assim, é possível ter uma melhor visão global da exposição dos sistemas e da informação da empresa, para uma melhor prevenção de possíveis problemas.

Para auxiliar a monitorização já existente do perímetro empresarial, foi desenvolvido o RoboSCOUT. O RoboSCOUT é uma ferramenta que permite a pesquisa, normalização, unificação, armazenamento e visualização de forma automática dos dados obtidos sobre sistemas vulneráveis e informação crítica exposta na Internet com recurso a plataformas OSINT.

Com a integração de duas fontes, o Shodan e o Censys, vão ser efetuados varrimentos periódicos à rede corporativa com o objetivo de encontrar vulnerabilidades específicas nos sistemas, a deteção de dispositivos vulneráveis e de camaras IP. Estes três pontos foram considerados como sendo os mais importantes para um auxílio da defesa do perímetro de sistemas. O uso de outras duas fontes, o Google e o Pastebin, vai permitir a pesquisa de informação crítica sobre a empresa que tenha sido exposta como contas de utilizadores ou informação de caráter confidencial que possa estar exposta na Internet.

Depois de recolhida a informação, o RoboSCOUT vai proceder a normalização dos dados com o uso de termos genéricos e agrupar num único relatório para depois de armazenado, ser consultado numa plataforma de visualização para permitir uma análise dos dados.

A implementação do RoboSCOUT foi executada na rede empresarial da Altice Portugal, analisando os sistemas presentes na mesma. Depois de implementado, foram efetuados testes, de acordo com os requisitos iniciais, para comprovar a eficácia do projeto que produziram os resultados expectáveis.

Palavras-chave: monitorização, inteligência, cibersegurança, ataque, automação

2021

Abstract

In the current days, the concern with cybersecurity has gained more importance than ever becoming one of the most important matters for any organization. Also, the use of remote work brought concerns to the definition of the companies' defense perimeter. The use of devices known as IoT also brought a security concern since most of them are not properly secured and can became an entry to nefarious agents. This perimeter is exponential to the size and diversity of the systems within in it, which may cause a challenge in securing it. In addition to the physical perimeter, there is also the intelligence perimeter. The latter is composed of crítical information that, if leaked, can cause the same amount of damage to the company as a flaw in the physical one. So, the risk associated to an attack is directly correlated with the protection given to those perimeters.

Knowing the exposure to the Internet that a company has through its perimeter can be challenging in terms of complexity since multiple tolls had to be used to cover all those cases and in terms of gathering the information since they would come in different forms and types. In addition, the manual search in those tolls can become very dispendious in terms of time and effort.

To help the already existing defense of the perimeter, we created RoboSCOUT, an automated search and process engine which uses OSINT mechanisms to search for the services that are exposed to the Internet and critical information that might have been leaked. This toll automatically coordinates the search, normalization, merging, storage, and visualization of the data obtained by those sources. The goal of this toll is to gather knowledge on what the company might have exposed and critical data that otherwise could be hard to find. Using two sources, Shodan and Censys, RoboSCOUT will search for specific vulnerabilities on the systems, detection of vulnerable devices, and IP cameras. The use of other two OSINT sources, Google and Pastebin, will be used to search for critical information that might have been exposed on the Internet such as user accounts or other type of confidential information. RoboSCOUT will then proceed to correcting and normalizing the data for merging it in one report for better comprehension of the current situation. This generated report will then be sent to a visualization platform that will show the data for better analysis.

The implementation of RoboSCOUT in Altice Portugal infrastructure was tested with use cases that, in accordance with the system requisites, proved the efficiency of the toll since they gave the expected results.

Keywords: monitoring, Intelligence, cybersecurity, attack, automation

2021

Conteúdo

Lista de Figuras

2021

Lista de Tabelas

Capítulo 1

Introdução

Com o aumento significativo de dados armazenados, a automatização de processos e a melhoria das práticas da segurança de informação são essenciais no mundo empresarial para garantir um funcionamento seguro e estável de todos os seus componentes.

Uma das melhores práticas é a monitorização contínua de todos os ativos da empresa pois só assim é que proactivamente, se conseguem encontrar potenciais vulnerabilidades que possam ser exploradas por atacantes.

A rede de uma organização é altamente heterogénea tendo todo o tipo de serviços e infraestruturas. Com o uso de múltiplas fontes, cada uma analisando um espetro de informação diferente, é possível obter uma área maior na pesquisa por vulnerabilidades no perímetro de defesa. A existência de uma ferramenta como o RoboSCOUT irá permitir uma pesquisa constante e automática pela exposição que os serviços tenham para a Internet assim como informação disposta online.

O RoboSCOUT vai auxiliar os diferentes mecanismos já existentes na organização como o BitSight [1], o Qualys [2] e o Cycognito [3] com o recurso a plataformas OSINT para cobrir o espetro de sistemas e informação crítica exposta na Internet. A adição desta ferramenta ao leque das fontes já usadas pela organização irá ser uma mais-valia, pois, a existência de fontes OSINT irá permitir à organização saber que dados dos seus sistemas estão disponíveis públicamente. Para além disso, a procura por informação crítica também é um novo espetro de análise que irá ser implementado, cobrindo assim cada vez mais possíveis pontos de ataque ou entrada na rede organizacional.

1.1 Motivação

Com a quantidade de informação atualmente presente, os atacantes conseguem recorrer a plataformas que permitem estudar os serviços duma forma trivial e obter informação sobre os sistemas organizacionais. Consoante o nível de conhecimento do atacante, os danos dos ataques podem afetar a operacionalidade da rede. Vetores de ataque que possam ser estudados com o uso dessas plataformas, que estão especificamente construídos para obter informações de sistemas e as suas vulnerabilidades, terão maiores probabilidades de sucesso.

"if you can't beat them, join them" [4]. A melhor defesa para este tipo de ataque é tornar uma rotina dos atacantes numa rotina da empresa, em que se irá proactivamente pesquisar por sistemas e serviços expostos na Internet. Para além da componente tecnológica, também vai ser pesquisada informação crítica exposta na Internet.

A defesa da infraestrutura exposta da empresa é importante e os mecanismos que aliem a pesquisa constante de possíveis vetores de ataque que surjam na rede, assim como a pesquisa de informações consideradas críticas para a empresa são uma mais-valia e uma forte adição ao leque já existente de ferramentas presentes na organização.

Portanto a existência de uma solução que seja complementar aos mecanismos defensivos já implementados de uma empresa e de uma forma *ad-hoc* (o uso do sistema quando necessário) permite o aumento do leque das ferramentas utilizadas.

1.2 Objetivos

Passando por uma solução complementar, o RoboSCOUT tem como objetivo ser utilizado de uma forma tática e focada na descoberta preliminar de indícios de exposição dos sistemas corporativos (de onde advém o SCOUT) tanto da Altice Portugal como qualquer cliente que queira efetuar essa descoberta.

As plataformas tradicionais e em uso têm uma larga janela temporal face a divulgação das vulnerabilidades com descrição e remediação dos sistemas analisados. Plataformas OSINT conseguem ter melhores condições para a deteção e análise dessas situações.

Para além da melhor eficácia na busca de vulnerabilidades um dos outros objetivos deste projeto passa por procurar dispositivos IoT de uma forma mais rotineira, que possam ser conectados, associando as marcas de fabricante e a lista providenciada dos IPs da empresa.

Em ambos os casos é necessário que a automação no uso destas fontes tenha uma gestão com eficácia e eficiência dos licenciamentos providenciados para o uso das ferramentas. Deve produzir os resultados esperados no menor período de tempo sem que isso comprometa as limitações dos licenciamentos (como o número de *queries* efetuadas por unidade de tempo).

Também um dos objetivos, que destaca a ferramenta, é a bpesquisa em sítios *web* do que é considerada informação crítica para a empresa. Através do uso de domínios corporativos, como terminologias nas contas de utilizador ou de correio eletrónico, obter informação que possa ter sido exposta na Internet e que contenha algumas das terminologias de ficheiros corporativos ou de contas corporativas.

Após as pesquisas nas várias ferramentas serem efetuadas com o uso do motor de pesquisa, os dados terão de ser normalizados e armazenados pelo motor de processamento para serem consultados em plataformas de visualização.

O objetivo final será a implementação da automatização dos sistemas de pesquisa e processamento com recurso a um motor de calendarização que irá coordenar as chamadas de cada um.

1.3 Organização do documento

Este relatório está organizado da seguinte forma:

- Capítulo 2 (Contexto) – É apresentado o ambiente em que o projeto se integra e tecnologias ou ferramentas já desenvolvidas e que serão utilizadas;
- Capítulo 3 (O Sistema) – É apresentada a arquitetura do sistema justificando as decisões tomadas ao longo do processo de planeamento do RoboSCOUT;
- Capítulo 4 (Implementação) – Neste capítulo são descritos em detalhe os métodos utilizados na implementação do sistema;
- Capítulo 5 (Testes e Resultados) – São apresentados os vários testes submetidos ao RoboSCOUT e os seus resultados;
- Capítulo 6 (Conclusão) – Neste capítulo é sumarizado todo o trabalho realizado na criação e implementação do RoboSCOUT onde são apresentadas as conclusões dos resultados que foram obtidos. Também são referidos pontos de trabalho futuro para a ferramenta.

Capítulo 2

Contexto

Este capítulo providencia um contexto sobre o problema que este projeto se debruça assim como áreas chave que são essenciais para a construção do mesmo. Vai ser explicado o conceito de OSINT com vários exemplos de plataformas e o tipo de informação que cada plataforma oferece e também a escala de classificação das vulnerabilidades encontradas.

2.1 OSINT

Open Source Intelligence [5] ou "Inteligência" com base em Fontes Abertas tem como foco a recolha de informações de fontes públicas neste caso, *websites*. O uso deste tipo de recolha de informação traz benefícios como o baixo custo de pesquisa já que a informação se encontra pública e em grande escala, a disponibilidade dos serviços que providenciam os resultados e uma elevada taxa de atualização dos sistemas.

2.1.1 Shodan

Shodan é um motor de pesquisa para analisar dispositivos na Internet criado em 2009 e que atualmente é o líder do mercado em motores de busca de dispositivos. Todos os tipos de dispositivos conseguem ser encontrados pelo Shodan desde camaras web, servidores ou frigoríficos.

Funcionamento do Shodan

O algoritmo de pesquisa do Shodan é secreto, no entanto, é sabido que os scanners do Shodan estão posicionados globalmente o que dá um resultado de pesquisa abrangente. A procura de dispositivos sendo distribuída pelos vários scanners, permite uma maior eficácia nos portos analisados e tipos de dispositivos encontrados, nomeadamente, dispositivos IoT.

Com o uso da interface web, é possível adicionar uma lista de IPs que se pretendam ser analisados onde é apresentado para cada um, os portos abertos e os serviços que estão a ser executados. No entanto, a execução e o tipo de filtros que podem ser utilizados são limitados, não podendo ser utilizada para os objetivos do projeto.

Na API, o resultado da pesquisa efetuada pode ser visto em formato RAW na interface da linha de comandos ou guardado para um ficheiro JSON que pode ser facilmente convertido em CSV através do uso de scripts (Ruby por exemplo).

2.1.2 Censys

O Censys [6] é um motor de busca de serviços criado em 2015 que tem como objetivo providenciar uma imagem quase em tempo real da Internet como um serviço público. É um sistema baseado em tecnologias como o ZMap [7], uma versão mais rápida que o Nmap [8] para percorrer os endereços IPs, e o ZGrab [9], uma vertente do ZMap que permite obter os *banners* dos sistemas.

O Censys utiliza um método de pesquisa conhecido como scan horizontal onde todos os endereços IP são analisados apenas num porto de cada vez e coleta o *banner* (meta data sobre o dispositivo ou dos serviços a executar nesse dispositivo) dos portos com resposta.

O Censys dedica certos scanners a portos específicos e possui outros que verificam múltiplos portos. O foco desses scanners dedicados é os protocolos HTTP/HTTPS. Os scans são calendarizados com base na popularidade de certos portos e redes no espaço de endereçamento IPv4 e todas as entradas na lista são analisadas para verificar a sua idade. Caso tenham mais de 24 horas, são novamente atualizadas fazendo com que o *data set* esteja sempre o mais atualizado possível.

2.1.3 ZoomEye

Tal como o Censys e o Shodan, o ZoomEye [10] é também um motor de busca de dispositivos expostos na Internet que através da análise de nós pelo motor presente no *backend*, as suas características são discriminadas e assim obtêm informações sobre os mesmos. Através de duas bibliotecas, Xmap [11] e Wmap [12], o ZoomEye consegue reconhecer componentes de websites (linguagem do server, aplicações web, …) e também através da integração do

Nmap, um *software* que realiza scan a portos, permite aferir que tipos de serviços estão presentes nos sistemas.

2.1.4 Google Dorking

O Google, um motor de busca amplamente conhecido, possui mais termos de pesquisa do que o utilizador normal utiliza, com o uso de filtros específicos. O uso desses filtros permite realizar pesquisas mais específicas como a procura de extensões de ficheiros específicas, palavras presentes no link ou no corpo do site, datas de publicação, entre outras.

2.1.5 Pastebin e GhostBin

Quando um atacante descobre uma vulnerabilidade ou um possível vetor de ataque e quer partilhar com alguém certamente irá partilhar de forma completamente anónima e que não seja possível rastrear o seu paradeiro. Websites como o Pastebin [13] ou Ghostbin [14] que inicialmente foram criados como um simples bloco de notas online, onde várias pessoas poderiam consultar ou editar informação, tornaram-se nos websites mais usados para partilha de informações críticas, scripts que permitem a exploração de vulnerabilidades ou qualquer tipo de informação que possa ser partilhada sem ser possível rastrear o autor original da públicação tornando o sistema anónimo.

2.1.6 SpiderFoot

SpiderFoot [15] é uma ferramenta de reconhecimento que junta mais de 100 fontes de dados de acesso público (OSINT) e que permite obter informações sobre o alvo requerido desde endereços IPs, nomes de domínios, endereços de e-mail, etc. Assim, o *SpiderFoot* consegue providenciar informações variadas sobre o alvo o que permite aferir a existência de potenciais fugas de dados ou vulnerabilidades presentes nos sistemas.

Domain Name	1	3	2020-06-26 01:24:16
IP Address	178	268	2020-06-26 01:22:46
IPv6 Address	24	24	2020-06-25 22:28:27
Internet Name	136	689	2020-06-26 01:20:39
Internet Name - Unresolved	12	50	2020-06-26 01:13:43
Netblock Membership	21	40	2020-06-26 01:17:59
Open TCP Port	277	503	2020-06-26 01:19:09

Figura 2.1- Resultado de um teste do SpiderFoot

2.2 Vulnerabilidades

2.2.1 CVE

O programa CVE (*Common Vulnerabilities and Exposures)*, lançado em 1999 e operado pela *MITRE*, é uma lista disponível públicamente com as diversas falhas de segurança existentes. A cada uma dessas falhas é atribuído um número único que a identifica inequivocamente.

Esse número único é uma combinação do ano em que foi descoberta e um número arbitrário entre 4 e 7 algarismos separados por um hífen. Para ser atribuído esse número tem de existir 3 critérios que a qualifiquem. A falha pode ser resolvida independentemente da existência de outros *bugs*, a falha tem de ser confirmada pelo fornecedor do serviço e documentada e só pode afetar uma base de código.

2.2.2 CVSS

Cada vulnerabilidade tem um impacto diferente dadas as suas características, o que pode provocar diferentes níveis de criticidade na organização. Para que exista um termo de comparação entre elas foi criado o *Common Vulnerability Scoring System*.

O CVSS é, atualmente, o standard para essa definição e calculo da criticidade utilizada pela indústria tecnológica. Com diferentes características e grupos de medida, é possível obter uma classificação para cada uma das vulnerabilidades. O CVSS, atualmente na versão 3.1, é composto por três grupos de métricas: base que é obrigatório para o cálculo da medida, e os outros grupos que são facultativos, temporal e de ambiente. O primeiro grupo define as características da vulnerabilidade como o grau de facilidade de exploração e o impacto que causa. O grupo temporal é dinâmico dada os diferentes casos que podem ocorrer com a

vulnerabilidade onde pode diminuir, por exemplo, se houver uma correção para a mesma. O grupo de ambiente representa as características relevantes e únicas num ambiente específico como medidas de segurança ou medidas de mitigação caso um ataque seja bem sucedido. Estas medidas são então agregadas e é produzida uma pontuação de 0.0 a 10.0 como mostra a tabela 2.1.

CVSS	Impacto
0.0	Nenhum
0.1-3.9	Baixo
4.0-6.9	Médio
7.0-8.9	Alto
9.0-10.0	Crítico

Tabela 2.1 - Pontuação CVSS

2.3 Princípios de um ataque

Com o aumento de ataques informáticos cada vez com mais complexos, torna-se crítico para a empresa um ataque destes ser bem-sucedido, pois pode pôr em causa o bom funcionamento da organização. Para além de melhorar as defesas existentes, também é importante reduzir ao máximo a superfície de ataque do adversário.

2.3.1 Vulnerabilidade de Camaras e IoT

Camaras IP podem instaladas com facilidade para qualquer motivo. Com o endereçamento publico da camara [16] os atacantes conseguem através desse endereço, caso esteja vulnerável, ligar-se à mesma conseguindo um ponto de entrada para a rede corporativa.

Essas vulnerabilidades são, por exemplo, o uso de credenciais de login de origem, palavras-passe triviais e mecanismos de cifra pobres ou inexistentes.

Existem múltiplos ataques que podem ser efetuados a estes tipos de dispositivos como o *eavesdrop* que consiste na monitorização de dados que passem sobre a camada de rede para obter o texto em claro das comunicações, o *Man in the Middle* onde é implementado um protocolo de SSL malicioso que permite o atacante redirecionar o tráfego para o seu computador e observar as comunicações em claro ou, então, utilizando analisadores de pacotes (*sniffers*) para entrar na rede como membro legítimo e controlar os dispositivos presentes naquela rede. [17]

2.3.2 Sistemas vulneráveis

A presença de serviços que possuam vulnerabilidades em sistemas empresariais podem facultar uma entrada na rede corporativa. Protocolos mal configurados, portos abertos ou atualizações que não tenham sido efetuadas para eliminar vulnerabilidades descobertas, são alguns dos vetores que um atacante pode ter para ganhar acesso sobre a rede empresarial.

Um porto aberto não significa perigo por estar aberto, mas sim por poder estar mal configurado, vulnerável ou com fracas implementações de segurança de rede. Existem portos que podem ser infetados com mais facilidade que outros com recurso a *worms* (*malware* replicativo) por exemplo ou portos que como não deveriam estar abertos não têm políticas de segurança associadas. Por porto aberto neste projeto define-se estes casos.

2.4 Pesquisa de informações críticas

A partilha indevida de informação privada corporativa como contas de utilizador pode acontecer com mais frequência dada a conjuntura atual. Se um atacante encontrar ficheiros que tenham credenciais de acesso ou outro tipo de informação que lhe consiga dar vantagem no seu ataque os danos provocados pelo mesmo podem ser maiores. Para além disso, o extravio dessa informação também provoca um problema de privacidade.

A pesquisa de informação crítica na Internet, com o uso do Google Dorking e do Pastebin, permite que essa informação relativa a empresa seja detetada duma forma automática e atempada, dando tempo para que sejam tomadas medidas de mitigação para os problemas encontrados.

Com o uso do BluePrism na pesquisa de informações através do Pastebin e do Google Dorking a pesquisa pode ser efetuada de forma completamente automatizada conseguindo os resultados no momento.

2.5 Conclusão

Este capítulo apresenta as fontes que existem na área de OSINTs a serem utilizadas no projeto. Também foi explicado as diferentes formas que um atacante pode ter para efetuar um ataque e a criticidade que apresentam informações que estejam públicamente disponíveis. Não

foram apresentados trabalhos relacionados já que não existe uma ferramenta que proporcione a mesma área de pesquisa que o RoboSCOUT.

No próximo capítulo será apresentada a arquitetura que foi planeada para a implementação do RoboSCOUT justificando as decisões tomadas.

Capítulo 3

O Sistema

Estudando as medidas implementadas atualmente na empresa, mais especificamente no CyberSOC, verificou-se que a vertente de análise de sistemas do RoboSCOUT irá ajudar as fontes já existentes para obter uma informação mais completa. A pesquisa de informação crítica exposta é uma nova vertente na prevenção de ocorrência de eventos de segurança em sistemas e serviços expostos na Internet, através da utilização da informação adquirida em fontes OSINT.

3.1 Requisitos de Sistema

O RoboSCOUT permite a adição de uma nova linha de prevenção a múltiplos vetores de ataque, tirando também partido da sua completa automatização.

O uso das fontes em separado e a sua respetiva análise iria tornar o método de análise, complexo e demorado. Por isso, é proposto o uso de um único motor de pesquisa que alie o uso de RPA, para a pesquisa na Internet de informações críticas, com o uso de APIs para analisar sistemas expostos da empresa e possíveis vulnerabilidades.

Sendo assim foi definido um conjunto de pressupostos que o sistema deve cumprir para ser possível a criação e integração com os diversos sistemas já presentes na infraestrutura interna da empresa. Os requisitos da solução são:

- A existência de modularidade, isto é, os motores de pesquisa serem independentes entre si e a existência do mesmo grau de abstração na ingestão dos dados por parte do motor de processamento. Isto permite adição ou remoção de motores de pesquisa sem corromper o sistema.

- Deve existir escalabilidade na análise de *assets* e na pesquisa de informação na Internet sem causar o bloqueio do sistema ou uma janela temporal desproporcionada.
- Existência de ficheiros globais que permitam uma abstração do uso da ferramenta. Ficheiros de âmbito permitem colocar os *IPs* que serão analisados pelos motores de busca e os ficheiros de contexto servirão para colocar a informação adicional que será necessária e específica para cada caso de uso.
- A capacidade de uso individual dos motores de busca existentes ou a execução manual (sem recurso a calendarização) para permitir alguma necessidade de verificar os dados em emergências como descoberta de novas vulnerabilidades ou de informação crítica exposta.

3.1.1 Escolha de Fontes

Com o leque presente de várias fontes na análise de sistemas e de pesquisa de informação a primeira escolha torna-se óbvia dado um dos objetivos principais do trabalho. O Shodan irá permitir uma análise de sistemas e serviços expostos, com retorno de informações importantes e específicas de cada ativo, assim como a pesquisa facilitada de dispositivos IoT. Para uma melhor cobertura da área de resultados, nomeadamente protocolos HTTP e HTTPS, o uso do motor de busca Censys irá permitir obter a imagem mais atualizada possível dos sistemas organizacionais. Sendo que cada um dos motores também difere nos portos mais comuns que pesquisa [18], como o Shodan que pesquisa mais frequentemente nos portos 80 e 1177 e o Censys que pesquisa nos portos 443 e 21, o uso de ambos irá resultar numa lista de resultados mais abrangente e completa.

Para a pesquisa de informação exposta na Internet torna-se óbvio o uso do maior motor de busca, o Google. O uso do motor vai permitir a pesquisa de ficheiros ou de texto contidos em sítios *web* de uma forma eficaz. Para além desse motor também irá ser utilizado como fonte um dos maiores sites de partilha de informação de informáticos ou *hackers*, Pastebin.

3.1.2 Casos de Uso

Para alinhar os requisitos do projeto foram criados casos de uso para especificar o comportamento dos módulos em que foram definidos o âmbito (uma lista de IPs da rede corporativa ou uma lista de domínios corporativos), o contexto (especificação de inputs que o módulo necessita para efetuar a pesquisa), o output expectável e as fontes necessárias para cada um dos casos de uso.

O primeiro caso de uso tem como base uma lista previamente definida de IPs de dispositivos empresariais que vão ser o alvo da pesquisa e um ficheiro com o identificador, de acordo com a base de dados do CVE [19], das vulnerabilidades que se querem pesquisar com uso do Shodan e do Censys.

O segundo caso de uso tem como objetivo a pesquisa de dispositivos que sejam considerados vulneráveis e que possam representar uma ameaça a estrutura da empresa seja pela existência de portos abertos indevidos, protocolos mal configurados ou por ativos que possam ter palavras-passe de acesso originais.

Um vetor de ataque que não existe proteção, a deteção de camaras ou dispositivos IoT que não tenham sido corretamente configurados e que estejam expostas é coberto pelo terceiro caso de uso. Para além do ficheiro de âmbito com os IPs corporativos, também será necessário um ficheiro de contexto que possua diversas marcas de fabricante de dispositivos IoT ou palavras-chave que sejam recorrentes nos *banners* deste tipo de dispositivos.

O ficheiro de contexto para cada um destes três casos varia conforme o ambiente de pesquisa e será necessário a existência de ficheiros com palavras-chave adequadas a cada caso (lista de dispositivos IoT, *passwords* originais de fábrica de cada serviço ou lista de identificadores de vulnerabilidades). O resultado providenciado destes três casos de uso irá ser uma lista de ativos com as informações necessárias para fazer uma análise posterior dos mesmos.

Existem mais dois casos de uso que recorrem às outras fontes referidas (Pastebin e Google Dorking) e que, como âmbito, vão ter os domínios corporativos para permitir a pesquisa de informação que possa ter sido exposta.

O quarto caso de uso vai ter como objetivo a pesquisa por palavras-chave, usando o ficheiro de contexto, que possam estar presentes em documentos disponíveis na Internet.

Para além de informação documental, a pesquisa de contas de colaboradores ou *passwords* de serviços que tenham sido expostas também é fulcral, e o quinto caso de uso faz essa pesquisa através de terminologias conhecidas das contas de utilizador.

Abaixo está a Tabela 3.1 com o resumo dos casos de uso.

#	Caso de Uso	Âmbito	Contexto	Fontes
1	Deteção de Vulnerabilidades	Lista de IPs	Lista de CVEs	Shodan, Censys
2	Disp Vulneráveis e PW origem	Lista de IPs	Palavras-Chave	Shodan, Censys
3	Deteção de Camaras	Lista de IPs	Palavras-Chave	Shodan, Censys
4	Informação Exposta	Lista de Domínios	*Strings*	Pastebin, Google Dorking
5	Contas Expostas	Lista de Domínios	Lista Domínios Corporativos	Pastebin, Google Dorking

Tabela 3.1-Casos de Uso

3.2 Estrutura do Sistema

A escolha destes motores de busca para o projeto foi analisada para cobrir o máximo de informação possível sobre os ativos e informação da empresa.

A informação obtida de ativos é mais que suficiente para permitir um relatório robusto e informativo e a adição de mais motores iria, pelo menos nesta fase, produzir informação redundante e que não iria trazer uma mais-valia em termos de esforço para o projeto. Para permitir uma categorização dos ativos da empresa e o quão crítico seria caso existissem vulnerabilidades ou pontos de entrada nesses intervalos de IPs, foi criado um sistema de níveis categorizado numa escala qualitativa de *low*, *medium*, *high* ou *crítical* na forma par chave-valor.

No caso de pesquisa de informações que possam estar expostas, os dois motores de busca escolhidos foram selecionados pela cobertura de informação que providenciam. Através de buscas personalizadas podem obter-se documentos com palavras chave que foram escolhidas previamente e serem processadas para posterior análise.

Para ser possível agregar todas as informações obtidas foi criado um repositório central para armazenar os resultados de pesquisa das fontes e que depois de serem normalizados e processados, são enviados para uma plataforma de visualização onde os analistas serão capazes de inferir informação de uma forma mais eficaz e atempada.

A arquitetura do RoboSCOUT irá ser dividida em 4 motores em que cada um tem a sua função:

- **Motor de calendarização –** Responsável pela execução automática dos módulos de pesquisa, processamento e envio para o repositório central.
- **Motor de pesquisa** – Responsável pela interação direta com as fontes e a recolha dos resultados.

- **Motor de processamento** – Responsável pela fusão dos resultados provenientes das 4 fontes, da normalização e cálculo de medidas de risco.
- **Motor de visualização** – Responsável pela interação com a informação previamente armazenada para que seja possível representar os dados a serem consultados.

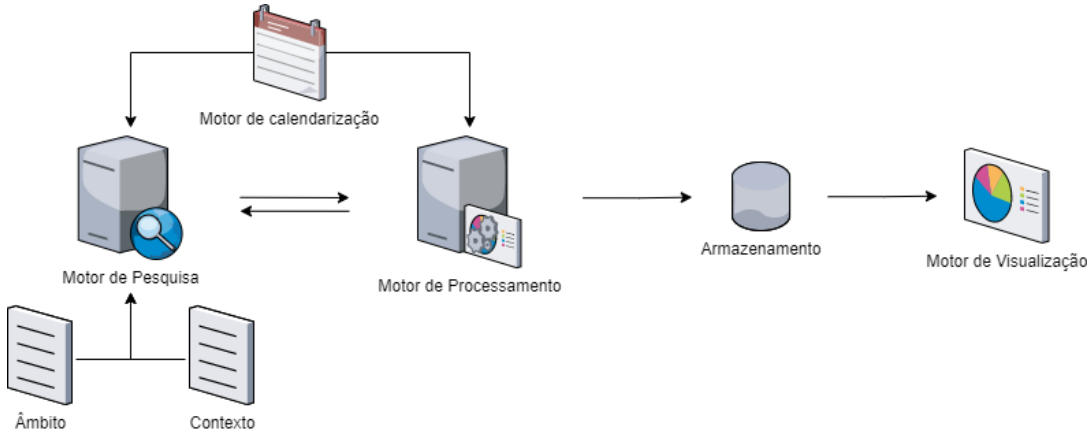


Figura 3.1- Arquitetura do RoboSCOUT

3.2.1 Motor de calendarização

O motor de calendarização é o responsável pela coordenação entre a execução dos módulos de pesquisa com recurso a um *Rakefile* [20], que irá ter as *tasks* necessárias para cada módulo e as subsequentes *actions* que são chamadas aos processos para encontrar e guardar a informação desejada. Esse *Rakefile* irá ser executado com recurso a *cronjobs* [21], uma biblioteca presente nos sistemas UNIX que permite executar comandos automaticamente numa hora previamente definida, já que a informação no Shodan e Censys é atualizada no mínimo, a cada 24 horas. Os motores que vão usar o RPA como meio de interação com a fonte, serão agendados pelo *scheduler* da ferramenta em si e não pelo motor de calendarização.

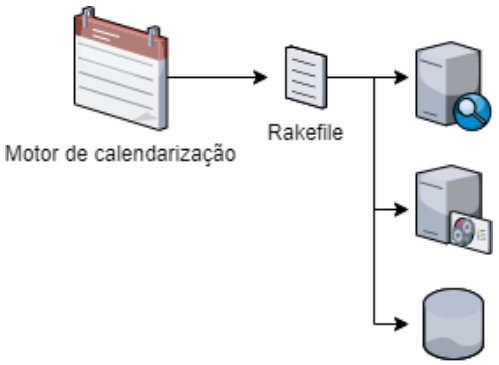


Figura 3.2-Motor de Calendarização

3.2.2 Módulos

Cada fonte irá ser um módulo, tanto para facilitar a adição de novas fontes como também para tornar a arquitetura do sistema mais linear. A figura 3.3 abaixo representa o mesmo.

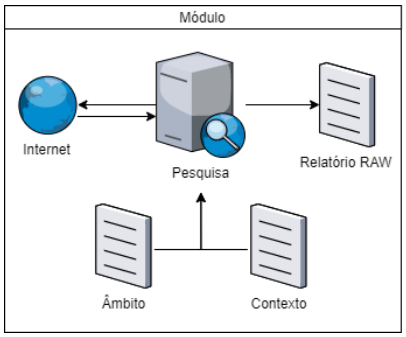


Figura 3.3 - Módulo de Pesquisa

Para que uma certa pesquisa possa ser efetuada, por exemplo, a pesquisa de dispositivos IoT na rede corporativa, é necessário fornecer como input ao script o âmbito (neste caso os IPs corporativos) e o contexto (*fingerprints* de dispositivos IoT como a marca do dispositivo).

3.2.3 Recolha de Informação

A primeira fase do RoboSCOUT é proceder à recolha de informação nas fontes previamente mencionadas. O processo de pesquisa de informação até ao seu armazenamento encontra-se delineado na figura 3.4. O motor de pesquisa vai ser o responsável pela pesquisa de informação em cada fonte e que depois irá ser entregue ao motor de processamento para o tratamento e normalização dos dados recolhidos.

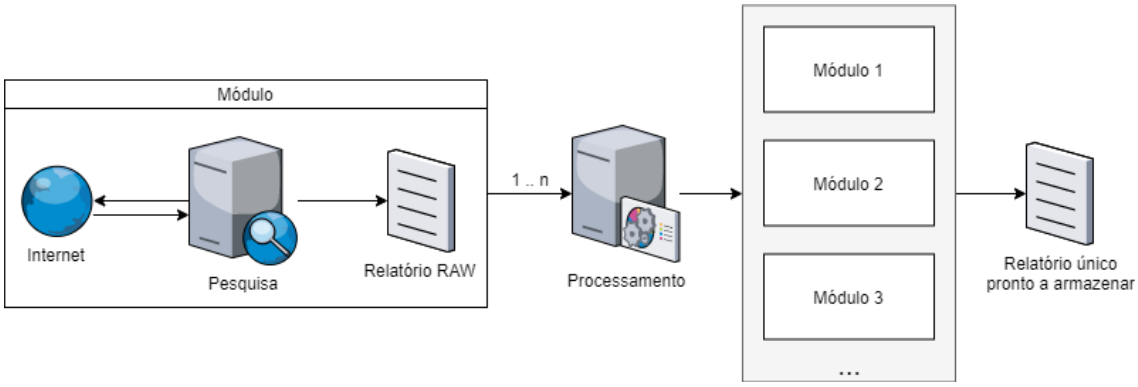


Figura 3.4-Arquitetura de recolha de informação

Sendo que cada fonte possui diferentes métodos de recolha, a existência de módulos que delimitem cada uma é essencial para permitir uma maior flexibilidade na adição de novas fontes, caso se justifique. Para duas das fontes (Google e Pastebin) foi definido o uso do BluePrism, uma ferramenta de RPA, que permite que robôs previamente programados repliquem interações humanas em páginas web e a extração dos resultados obtidos. Como esta tecnologia e as páginas web são dinâmicas, o processo de recolha terá de ser distinto para cada uma das fontes já que os formatos de saída dos dados recolhidos serão diferentes. Depois de serem recolhidos, irão ser enviados para o motor de processamento.

Para permitir um melhor desempenho do sistema foi decidido que para o Shodan [22], vai ser utilizada a CLI que é facultada, interligada com a linguagem de programação Ruby, podendo assim criar um ficheiro que vai fazer chamadas de sistema simulando a linha de comandos do sistema operativo e executar comandos com pedidos que irão permitir aferir a existência de ativos que possam estar vulneráveis e expostos na Internet. Para o motor de busca Censys vão ser usados pedidos HTTP também com recurso de um ficheiro Ruby.

Após ser efetuada a recolha de informação, como algumas das entradas recolhidas pelo Shodan vêm em formatos JSON incorretos os dados irão ser normalizados e corrigidos, para permitir o seu armazenamento no índice de *ElasticSearch* [23].

Agregação de Relatórios de Pesquisa

Depois da pesquisa ser efetuada e toda informação necessária for recolhida, o motor de processamento tem como função transformar a tabela JSON numa tabela CSV normalizada, para permitir a agregação de todas as pesquisas dos vários casos de uso referentes a cada método de pesquisa (CLI e RPA) já que a informação obtida pelo Shodan, por exemplo, não tem pontos em comum com informação retirada da Internet.

Por isso, foi criada uma tabela única com os vários pontos que são necessários para colocar a informação proveniente de cada de uso e em que, em cada linha para efeitos de organização dos dados, é acrescentado qual o caso de uso que a mesma se refere. Os cabeçalhos das colunas são os seguintes:

- mdav – motor de busca utilizado
- usecase – caso de uso referente à entrada
- ID – identificação única de cada entrada
- storage_date – Data de armazenamento do ativo na base de dados
- org – a organização associada ao IP
- isp – o fornecedor de serviço associado ao IP
- transport – tipo de protocolo de transporte (TCP, HTTPS, …)
- cpe – serviço responsável pela comunicação
- data – dados obtidos do banner
- asn – identificador do sistema autónomo
- port – porto utilizado na comunicação
- hostnames – nomes de domínio
- location – localização
- scan_time – data de entrada nos registos da fonte ou de pesquisa caso seja RPA
- http_tittle – caso seja website o url associado
- vulns – vulnerabilidades associadas ao ativo caso existam
- cert_valid – caso o serviço possua um certificado digital, dita se é valido ou não
- cert_validity – a data de validade do certificado
- overall_risk – risco calculado com as vulnerabilidades associadas
- ip_str – IP do ativo
- header– na pesquisa por RPA, é a pesquisa a ser efetuada pelas palavras-chave fornecidas
- word_hit – o resultado integral da pesquisa acima descrita

3.2.4 Visualização

O motor de visualização (figura 3.5) é o responsável pela criação de dashboards que permitem uma consulta sobre os resultados obtidos de uma forma intuitiva e eficaz através da ligação do *ElasticSearch* com a interface gráfica, *Kibana* [24].

Figura 3.5- Arquitetura do motor de visualização

3.3 Conclusão

Este capítulo aborda a estrutura e arquitetura planeada para o projeto assim como os requisitos do mesmo. Inicialmente existiam apenas duas fontes para integrar na arquitetura, mas ao longo do processo de desenvolvimento foi pensado a adição de mais duas fontes de informação (Pastebin e Censys) para permitir uma área de pesquisa e de obtenção de informação mais abrangente o que necessitou de uma reformulação da arquitetura do motor de processamento.

No próximo capítulo serão descritos mais detalhadamente os processos existentes no RoboSCOUT assim como as justificações das decisões tomadas.

Capítulo 4

Implementação

Neste capítulo é apresentada a descrição detalhada da implementação do projeto *RoboSCOUT*, os problemas que surgiram no desenvolvimento, as soluções desenvolvidas para os mesmos e as decisões tomadas ao longo do projeto. Como já referido, a linguagem de programação escolhida para este projeto foi *Ruby* já que é a linguagem de eleição para os projetos na *devOps* da empresa. Também foi necessária uma biblioteca de ferramentas extra para complementar o projeto tal como o *BluePrism*, para desenvolvimento dos processos que necessitavam de um método dinâmico de recolha de informação e também ferramentas pertencentes ao *stack* do *ElasticSearch* para armazenamento e visualização dos resultados. Para além disso também foi necessário o uso de uma máquina baseada em Linux onde a conexão é efetuada por SSH para permitir a execução dos processos e a sua calendarização.

Os quatro pilares para o projeto são os motores de calendarização, pesquisa, processamento e visualização.

4.1 Motor de Calendarização

A "chave" para o RoboSCOUT está no motor de calendarização pois é o que permite a execução automática e autónoma da ferramenta não necessitando assim, de intervenção humana para a sua execução.

Para permitir um melhor funcionamento da ferramenta foi decidido que a sua execução deveria ser diária e no período matinal para permitir depois a sua análise durante o dia.

4.1.1 Cronjobs

O uso de *cron jobs*, uma biblioteca presente nos sistemas UNIX, foi a base para este motor pois permite um agendamento de tarefas e comandos só sendo necessário a configuração

da *timeline* da execução. Para cada chamada de sistema é necessário a introdução da janela temporal que se pretende para a execução da mesma. Para além da hora da chamada, também é possível executar as chamadas certos dias da semana ou do mês. Após isso é necessário colocar o diretório onde a chamada irá ser efetuada e o comando em si que se quer executar.

Abaixo, são apresentados os comandos que irão ser agendados na máquina para a execução dos vários motores. A especificação do comando é a seguinte:

```
.---------------- minuto (0 - 59)
|   .------------- hora (0 - 23)
|   |  .---------- dia do mês (1 - 31)
|   |  |  .------- mês (1 - 12)
|   |  |  |  .---- dia da semana (0 - 6)
|   |  |  |  |
*  *  *  *  * diretório  comando
```

```
0 7 * * * RoboSCOUT/. rake sources:pull:shodan[‘altice’]
30 7 * * 1,3,5 RoboSCOUT/. rake sources:pull:censys[‘altice’]
0 8 * * * RoboSCOUT/. rake sources:pull:pastebin[‘altice’]
30 8 * * * RoboSCOUT/. rake sources:pull:dorks[‘altice’]
0 9 * * * RoboSCOUT/. rake process[‘altice’]
30 9 * * * RoboSCOUT/. rake elasticsearch:push[‘altice’]
```

4.1.2 Bundler e Rakefile

O *Bundler* [25] permite que se consiga compilar o RoboSCOUT. Para isso, o primeiro passo é realizar o *build* (bundle install) do ficheiro principal, que incluí as dependências das bibliotecas necessárias ao projeto, incluí-las na pasta do projeto e também carregar todos os ficheiros que são necessários a estrutura do mesmo criando, assim, um módulo do projeto em que cada ficheiro representa uma classe.

Após o módulo ser criado, o ficheiro *tasks.rake* permite criar um objeto da classe que irá ser executada. A figura 4.1 mostra uma das tarefas (*task*), a criação do objeto e a chamada dos métodos do mesmo onde a *task* é seguida pelo nome da mesma e os argumentos necessários na

chamada (nome da empresa). De seguida é criado o objeto Shodan e é chamado o método da classe.

```
task :case1, [:company, :dotenv] do |t, args|
  report = RoboScout::Shodan.new(args[:company])
  report.case1

end
```

Figura 4.1 - Tarefa do Rakefile

Com todas as tarefas criadas é possível assim executar os métodos sem a necessidade de criar um ficheiro só para esse efeito. Também existe a possibilidade de executar os comandos do ficheiro de tarefas manualmente e individualmente como por exemplo, executar apenas um caso de uso de uma das fontes, processá-la e enviá-la para visualização. A figura 4.2 ilustra a visualização global e hierarquia de ficheiros do motor de calendarização.

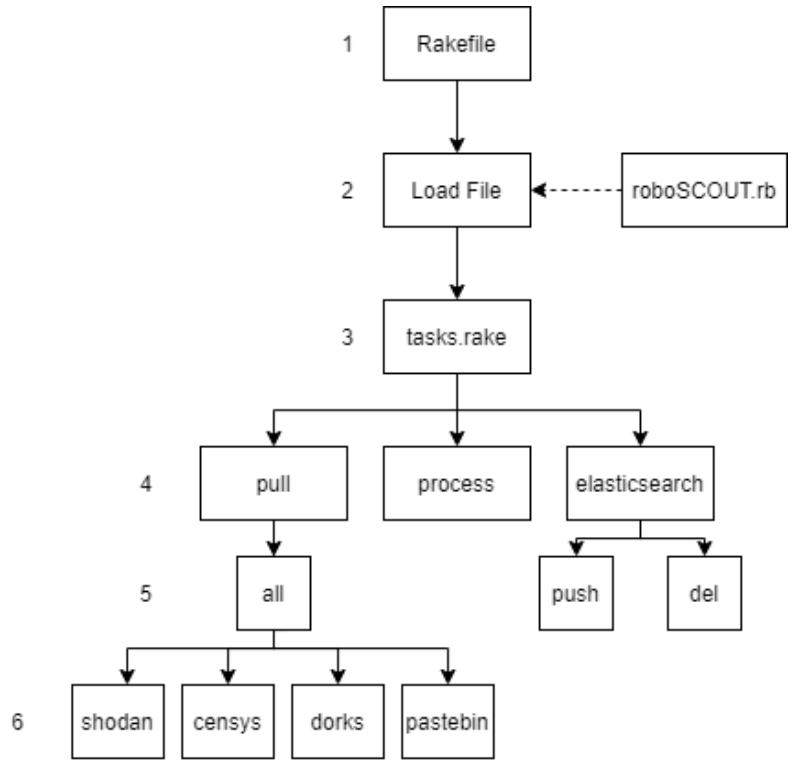


Figura 4.2 - Estrutura do Rakefile

1) Responsável pela agregação das diferentes tasks
2) Ficheiro que dita as bibliotecas e classes necessárias para a compilação do projeto
3) Ficheiro com as tarefas
4) Execução do motor de pesquisa, processamento e visualização
5) Tarefas auxiliares para os motores

6) Tarefas para permitir a execução de cada fonte individualmente

4.2 Motor de Pesquisa

Aquando da chamada do motor de calendarização, o motor de pesquisa é o responsável pela execução de cada uma das fontes e de guardar os resultados obtidos pelas mesmas. O seu formato modular permite uma separação lógica na pesquisa em cada fonte tornando a edição do programa ou adição/remoção de funcionalidades mais prática. A adição de novas fontes torna-se relativamente fácil já que é só respeitar o formato de armazenamento local dos dados e mudar algumas configurações no motor de processamento.

4.2.1 Pesquisa e extração no Shodan e Censys

Depois de criar uma conta no website do Shodan, retirou-se a chave da conta para poder inicializar o processo de comunicação com o recurso do CLI providenciado. Para cada caso de uso, a informação a ser colocada na interrogação ao Shodan como filtro difere como por exemplo, na pesquisa de dispositivos que possuam vulnerabilidades, onde é utilizado o CVE da vulnerabilidade com o filtro *vuln.* A figura 4.3 ilustra o processo de extração da informação necessária da fonte. Essa informação é devolvida num ficheiro comprimido e que contêm os resultados da pesquisa em formato JSON onde o mesmo é extraído e armazenado localmente para depois ser processado. A ordem de trabalhos do processo de extração é a seguinte:

1. É inicializado o cliente Shodan com a chave que foi obtida e que está guardada no ficheiro de configuração.
2. É feito uma interrogação ao sistema, dependendo do caso de uso em curso, sobre a existência de ativos que estejam dentro dos filtros a pesquisar para evitar créditos desperdiçados.
3. Caso existam ativos que se adequem aos filtros atuais, é feita a interrogação de forma a obter todas as informações sobre os mesmos onde de seguida é extraído o ficheiro dos resultados e guardado localmente.

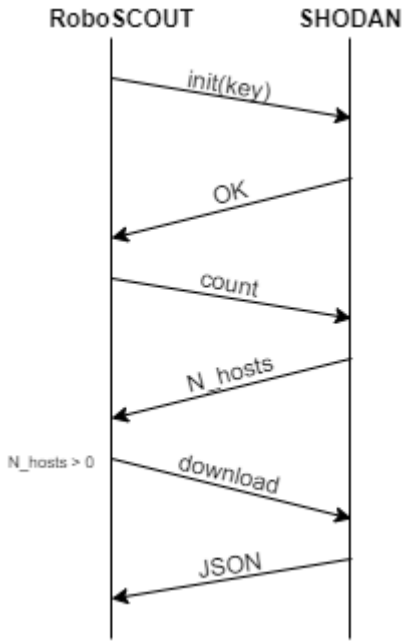


Figura 4.3-Extração de informação com Shodan

A API do Censys não é numa vertente de CLI, mas sim REST onde é necessário criar um pedido HTTP com a informação necessária como as credenciais de autenticação e a pesquisa a ser efetuada. Primeiramente é necessário realizar um pedido POST onde dada uma autenticação bem-sucedida, é obtida a informação simplificada da pesquisa. De seguida, é realizado um pedido GET onde se obtém a informação detalhada de cada resultado obtido e que após concluído é guardado em formato JSON localmente para processamento. A figura 4.4 ilustra o processo de comunicação em que os passos do processo são:

1. Com a autenticação proveniente do ficheiro de configuração e uma lista de endereços (ficheiro de âmbito) presente no corpo da mensagem é feito um pedido POST ao servidor no qual se obtém uma resposta com os serviços encontrados.
2. Para cada serviço encontrado é necessário efetuar um pedido GET para obter as informações mais detalhadas de cada entrada da lista e a cada iteração é adicionado ao ficheiro que existe localmente.

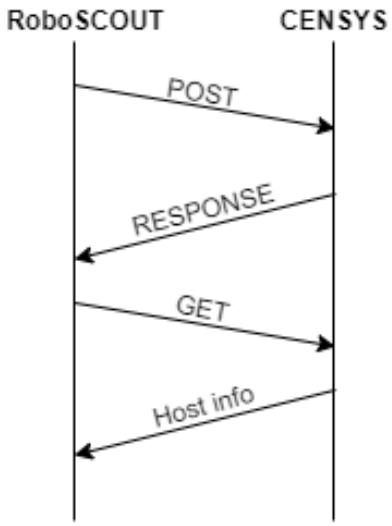


Figura 4.4- Extração de informação com Censys

4.2.2 Pesquisa e extração no Google Dorking e Pastebin

Outra vertente do projeto, para além da análise dos endereços corporativos é a pesquisa de informação crítica que possa ser encontrada online. Como a informação pode estar indexada em qualquer parte da Internet é necessária uma fonte que permita uma pesquisa em vários sítios online.

A pesquisa a ser efetuada usa recursos da própria pesquisa do Google que permite pesquisar por certas palavras no URL, no texto da página ou por extensões de ficheiros. O uso do método intext permite encontrar páginas que contenham as palavras escolhidas. O carater '*' funciona como uma wild-card podendo ser utilizado com outras palavras para encapsular a pesquisa como por exemplo "telecom.pt * pw". O uso destas queries é o que permite manusear de todas as formas possíveis as pesquisas consoante o requerido na altura.

Inicialmente o método de pesquisa com recurso aos Google Dorking foi desenhado com recurso a RPA. Isto quer dizer que a ferramenta iria escrever na barra de pesquisa, clicar pesquisar e recolher toda a informação presente na página de pesquisa como o url, título e resumo. Dada a volatilidade presente no código HTML da página, uma pesquisa que poderia ter um certo caminho definido até, por exemplo, ao url seria diferente noutra pesquisa efetuada e dado o carater estático da ferramenta de RPA não seria possível construir um método de pesquisa eficaz. Portanto foi utilizado uma API do Google que é utilizada para construir motores de busca internos em websites pessoais ou privados mais ainda com recurso ao RPA para efetuar os pedidos de forma mais linear e que não comprometesse a arquitetura já

planeada. Sendo assim, a figura 4.5 demonstra os vários passos efetuados pela ferramenta de RPA na pesquisa e obtenção de resultados:

1. Através de um ficheiro de configuração guardado na rede interna da empresa, são colocados os termos de pesquisa que deverão ser efetuados.
2. Com um ciclo que irá iterar cada termo de pesquisa, será utilizada a API do motor de busca Google para pesquisar e retornar os resultados em formato JSON e que irão ser colocados numa tabela única para posterior processamento.
3. Depois da tabela ser guardada localmente irá ser enviada para a máquina do motor de processamento, para os dados poderem ser trabalhados e selecionados.

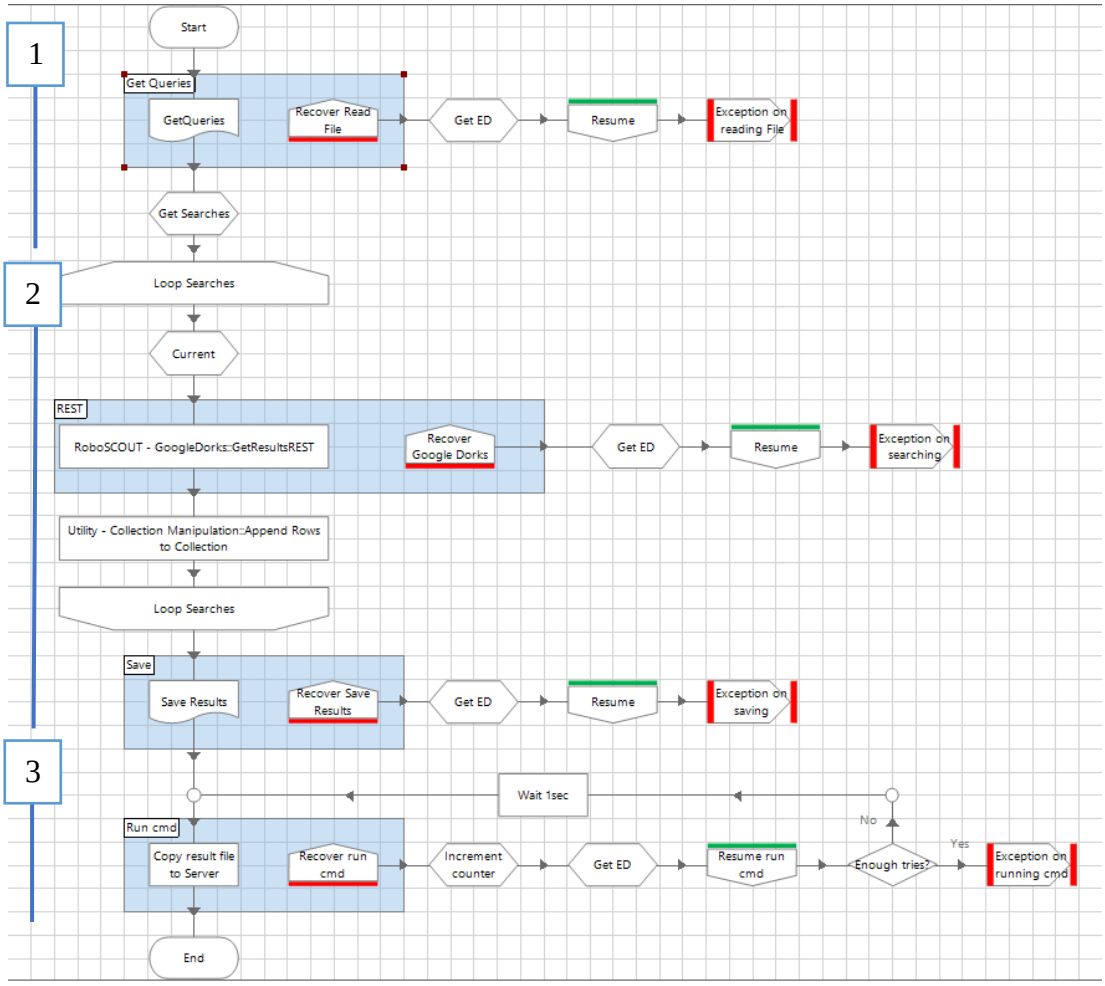


Figura 4.5-Extração de informação com Google Dorking

Para a fonte Pastebin onde o objetivo também é a pesquisa de informação crítica como contas de utilizador ou emails, são estudadas as cem pastas mais recentes presentes na fonte. Por pasta, entende-se um documento criado por uma conta (anónima ou não) e que possui um título, conteúdo, linguagem de programação (opcional) e o nome de autor caso não seja anónimo.

Com o recurso ao RPA as pastas são percorridas iterativamente onde, para evitar casos de duplicados com a atualização da lista, foi utilizada uma base de dados relacional que permite guardar o título e o corpo da pasta para verificar se a pasta já foi percorrida ou não. Caso não tenha sido, é adicionado a um ficheiro de formato JSON toda a informação relativa a pasta atual. Depois de todas as pastas percorridas, o ficheiro é enviado para o motor de processamento.

4.3 Motor de Processamento

A próxima fase do projeto é o processamento dos dados recolhidos. O motor de processamento é o responsável pela extração dos dados guardados localmente por cada fonte, pela sua normalização, cálculo de medidas e a agregação dos dados num relatório único. O motor de processamento também segue a modularidade que os motores de pesquisa possuem e cada interação do mesmo com cada relatório de cada fonte é feito de forma independente sendo possível retirar ou adicionar novas funcionalidades sem comprometer o bom funcionamento do sistema. Com a adição de uma fonte só é necessário a criação de um novo módulo e respeitar os formatos de dados (JSON) para que o motor de processamento seja capaz de ingerir os dados.

O primeiro passo do motor é a leitura linha a linha dos relatórios originais provenientes das diferentes fontes em que cada linha corresponde à informação detetada dado o caso de uso. Os relatórios originais estão separados por pastas em que cada uma corresponde a um caso de uso e que de forma iterativa são analisadas. Já sabendo onde cada relatório se encontra pela sua pasta, o motor de processamento tem os diferentes mecanismos de normalização da informação onde é identificado os atributos necessários e é feita a associação à categoria correspondente no relatório único. Para além disso, também é colocado o caso de uso em cada linha para melhor análise e filtragem de dados.

Como a informação dos ativos detetados estão em formato de texto no JSON é necessário transformar os dados que sejam necessários para números, datas ou coleções. Para as datas de deteção do ativo por parte das fontes é seguido o formato ISO 8061 [26] , ficando o standard de representação de data e hora. Caso existam vulnerabilidades nos ativos detetados é construída uma coleção que engloba o identificador da vulnerabilidade e a sua pontuação para ser calculado o risco.

Para o estudo da informação proveniente do Pastebin é procurado no texto as palavras chave armazenadas no ficheiro de contexto e que caso tenham a palavra mantêm-se.

4.3.1 Cálculo do Risco dos ativos

O ficheiro de âmbito, um dos ficheiros essenciais ao funcionamento do motor de pesquisa, também é necessário para o motor de processamento calcular o risco dos ativos que possuem vulnerabilidades já que possui o endereço IP associado ao ativo e o grau de criticidade.

Para que exista uma medida quantitativa que permita avaliar o grau de risco dos ativos que possuam vulnerabilidades foi criada uma matriz de risco que tem como eixos:

1) o grau de criticidade estático que é colocado no ficheiro de âmbito juntamente com o intervalo de IPs,

2) o impacto da vulnerabilidade de acordo com o standard CVSS 3.1 [27].

Aquando da deteção de uma vulnerabilidade num ativo o identificador da mesma é pesquisado na base de dados NVD [28] para obter a pontuação quantitativa e convertida para uma escala normalizada. Caso o serviço possua mais que uma vulnerabilidade, a pontuação escolhida é a correspondente a vulnerabilidade com o valor mais elevado na coleção existente. A figura 4.6 ilustra a normalização das escalas que depois serão utilizadas para calcular o risco final.

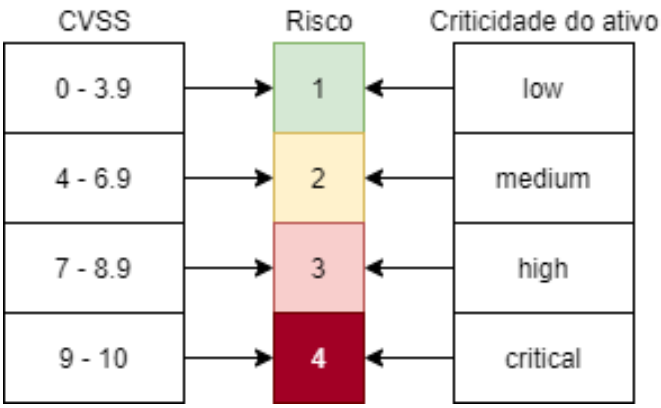


Figura 4.6-Normalização dos valores de um ativo

A figura 4.7 representa a matriz de risco utilizada onde o resultado da multiplicação entre os dois eixos será o valor final colocado na entrada desse ativo como risco global.

ATIVO \ CVEE	1	2	3	4
1	1	2	3	4
2	2	4	6	8
3	3	6	9	12
4	4	8	12	16

Figura 4.7 - Matriz de Risco

4.3.2 Relatório único de agregação

Com o uso de diversas fontes é natural que os dados obtidos venham em diferentes formatos o que implica que exista uma abstração da ferramenta perante cada um dos relatórios obtidos. Essa abstração começa numa tabela onde existem as colunas de informação consideradas necessárias e combinadas entre todas as fontes. Os termos escolhidos também foram pensados para que a adição de novas fontes possam utilizar o formato já construído. Este relatório após ser preenchido será enviado para o motor de visualização para poder ser analisado.

Grande parte dos termos foram adotados do Shodan já que é a fonte que mais informação obtém. Estes termos permitem aferir todo o tipo de informação sobre o ativo detetado e para além desses campos, foram adicionados os necessários aos outros casos de uso como os que irão recorrer ao método RPA, em que a informação obtida é de diferentes áreas pois corresponde a texto que tenha sido encontrado.

O relatório é inicializado e todas as informações são enviadas para a base de dados diariamente. O tipo de ficheiro escolhido foi o CSV já que é o que permite a maior facilidade de edição de dados e também porque torna mais fácil o envio para a base de dados. A figura 4.8 representa a estrutura do relatório onde está agregada a informação de diferentes fontes e que é agregada por uma única nomenclatura. A primeira linha corresponde aos termos escolhidos e as linhas subsequentes a cada ativo detetado ou informação encontrada.

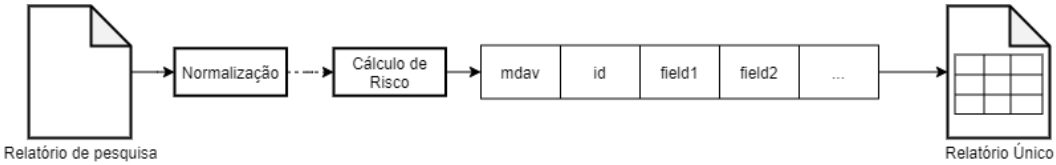


Figura 4.8 - Processamento de um relatório de pesquisa de um MDAV

4.3.3 Armazenamento na base de dados

Após o processamento da informação, o motor de processamento guarda localmente o relatório único com todas as entradas para cópia e é enviado para o cluster de *Elasticsearch*. Para isso, é efetuada uma ligação com a API e realizada uma conversão de alguns tipos de dados já que o envio é realizado com recurso a uma entrada em JSON. Quando a entrada é enviada para o cluster é necessário associá-la a um índice que é um mecanismo de organização em base de dados não relacionais. A existência de dois índices, o atual (*qa.roboscout.events*) e o histórico (*qa.roboscout.history*) permite transferir a informação do índice atual, onde se encontra a informação mais atualizada, para o histórico para permitir uma auditoria nas entradas.

4.4 Motor de Visualização

O quarto e último componente da ferramenta é o motor de visualização, o responsável pela criação de processos de visualização dos resultados de uma forma interativa, produzindo relatórios operacionais que permitam aferir conclusões sobre os dados de uma forma mais rápida e eficaz quando necessário.

Os dados que se encontram armazenados no índice qa.roboscout.events (Figura 4.9) permitem criar várias *visualizations* num único *dashboard* que forma o relatório operacional. Uma *visualization* permite obter uma representação visual dos dados que foram armazenados nos índices *Elasticsearch* e que podem ser tabelas, gráficos, histogramas ou medidores. Com o uso destas *visualizations* as *queries* executadas sobre os dados ficam guardadas automaticamente podendo aplicar filtros dinâmicos que alteram os dados apresentados.

Figura 4.9 - Resultados presentes no índice de ElasticSearch do RoboSCOUT

A criação de *visualizations* informativas permitem a análise geral dos ativos expostos da organização. A utilização desta ferramenta tem como base a criação de métodos de consulta simples e diretos que permitem visualizar a informação mais importante mais rapidamente. A necessidade de representar os dados em diversas formas permite uma heterogeneidade na visualização.

Caso haja necessidade de averiguar em mais profundidade os dados também é possível consultar diretamente no *dashboard* pelo ativo em concreto e obter os dados em forma bruta. Em anexo deste documento, Apêndice A, encontra-se o *dashboard* com todas as medidas e gráficos que foram implementados para uma melhor compreensão do que o motor de visualização produz.

4.5 Implementação do *software*

A solução para todo o projeto foi construída de raiz e alinhada com os objetivos e requisitos do mesmo e também tendo em conta o ambiente operacional da empresa. A figura 4.10 ilustra o diagrama de classes que foi desenhado para a construção da ferramenta e para facilitar a compreensão da estrutura do *software*.

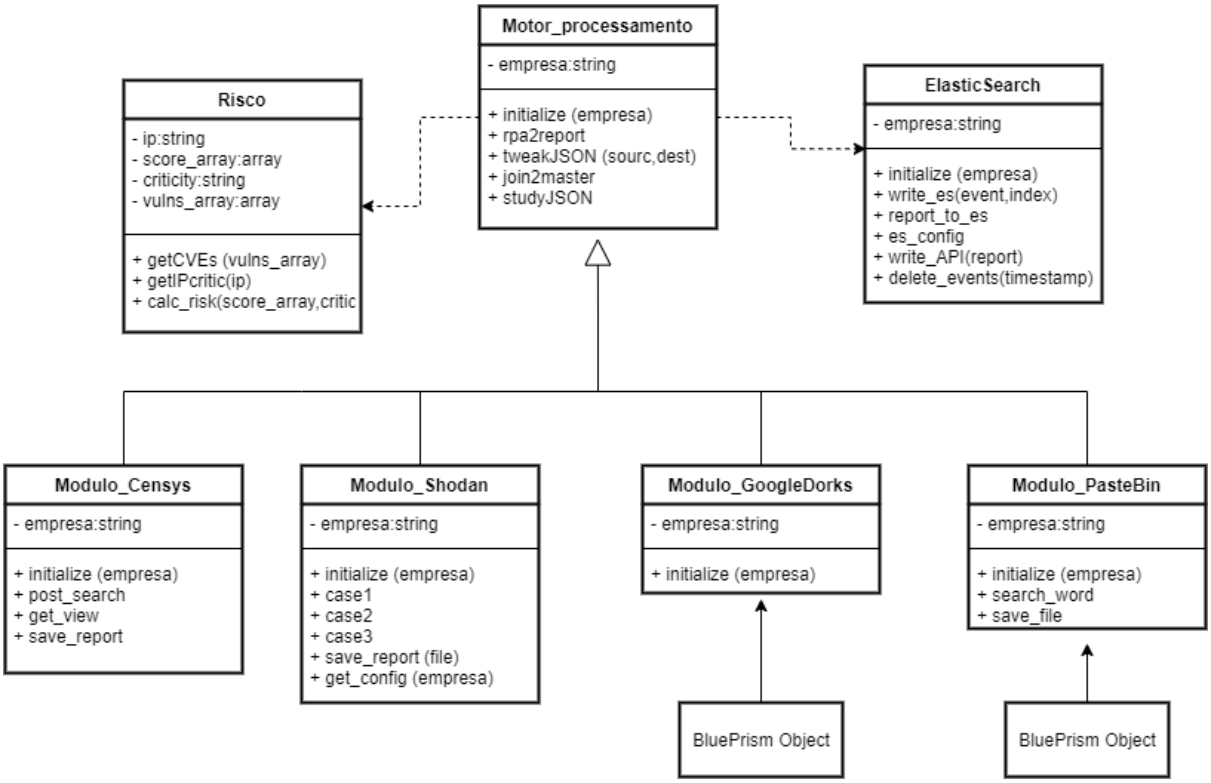


Figura 4.10 - Diagrama de Classes do RoboSCOUT

Classes do diagrama:

- **Módulo –** Responsável pela comunicação com cada fonte e a extração da informação necessária. As fontes que tem como base o uso da ferramenta de RPA necessitam que o objeto envie os relatórios para depois serem normalizados.
- **Motor Processamento** – Classe responsável pela extração de informação dos relatórios armazenados localmente para que seja processada e agregada ao relatório único. Também adiciona medidas calculadas durante o processamento ao relatório.
- **Risco** – Classe responsável pelo cálculo do risco. Contem métodos de pesquisa na fonte de vulnerabilidades e do grau de cada intervalo de IP para proceder ao cálculo. Depois é devolvido ao motor de processamento para que seja integrado no relatório único.
- **ElasticSearch** – Classe responsável pelo manuseamento da informação processada e do seu envio para a base de dados. Possui métodos para a eliminação de informação redundante e da transferência de informação do índice principal para o índice de histórico.

4.6 Conclusão

Este capítulo serviu para descrever todos os detalhes de implementação do RoboSCOUT no ambiente empresarial. Estes detalhes foram essenciais para que o projeto cumprisse todos os requisitos que foram definidos inicialmente e para um uso do sistema no quotidiano. Este capítulo também serve de meio de consulta já que foram apresentados alguns constituintes do relatório único e do *workflow* do sistema.

No próximo capítulo irão se descrever testes executados à ferramenta e os seus resultados para averiguar a sua eficácia e o cumprimento dos requisitos.

Capítulo 5

Testes e Resultados

Neste capítulo é realizada a testagem e avaliação do sistema RoboSCOUT. Com esta análise vai-se verificar se os casos de uso estabelecidos inicialmente foram alcançados (Capítulo 3.1.1). Para isso, foi feita uma *checklist* para os casos de uso apresentados anteriormente e feita a verificação de cada um. Dado que alguns casos de uso pesquisam, por exemplo, por dispositivos vulneráveis e não existem à data da pesquisa ativos que se encaixem nos filtros existentes, não se pode concluir efetivamente a eficácia dentro do perímetro da empresa já que não se irá colocar um ativo propositadamente vulnerável. No entanto, vai ser verificada a eficácia dos filtros ao pesquisar na rede global e certificar que de facto existem entradas na tabela.

5.1 *Checklist* dos casos de uso

Existem cinco casos de uso que foram marcados como fulcrais para o objetivo principal desta ferramenta. Para efeitos de teste, o ficheiro de âmbito, que serve como um dos *inputs*, foi preenchido com alguns dos intervalos de IPs disponíveis na empresa. Os intervalos correspondiam a servidores *web*, de correio e de serviços. O ficheiro de contexto foi preenchido com algumas *queries* correspondentes a cada caso de uso para verificar a sua eficácia na pesquisa. Para facilitar a testagem, cada caso de uso possui uma escala de 3 pontos, que irá ser apresentada após a análise de cada um: 1 que corresponde a falha do teste, 2 que comprova que o caso de uso funciona em termos de pesquisa global, mas que não foi testado no âmbito empresarial e 3 que declara o sucesso do caso de uso. De seguida vão ser apresentados os resultados dos testes.

No *dashboard*, para uma melhor filtragem e compreensão dos resultados foi colocada uma lista dos casos de uso que permite a escolha individual de cada um como ilustra a figura 5.1.

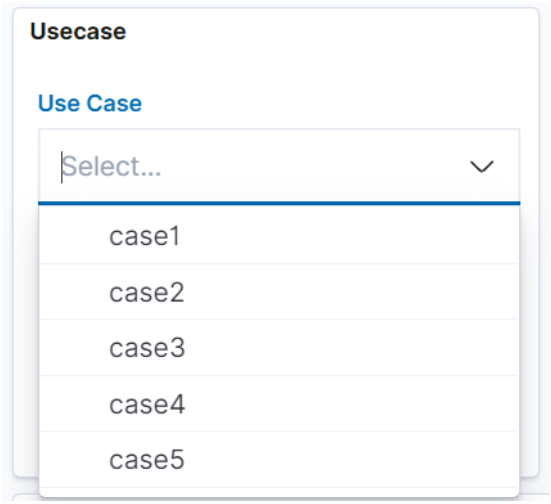


Figura 5.1 - Seleção do caso de uso

Deteção de vulnerabilidades específicas em sistemas

A existência de vulnerabilidades num sistema duma escala empresarial apresenta graves riscos ao seu funcionamento, e com uma grande área de exposição da rede a sua deteção e controlo é de difícil gestão.

Para permitir uma melhor gestão desses sistemas é relevante ter um mecanismo que permita a pesquisa direta de uma vulnerabilidade descoberta e realizar uma auditoria aos sistemas em busca de algum que possa ter sido afetado. Para isso, o primeiro caso de uso foca-se na descoberta desses mesmos sistemas que possuam vulnerabilidades descobertas pelos *crawlers* do Shodan ao fazer o confronto entre versões de ferramentas e as vulnerabilidades associadas a cada uma.

O caso de uso:

- **Âmbito –** Lista de IPs a pesquisar

```
○ 213.13.xxx.xxx/22: high
○ 213.13.xxx.xxx/23: low
○ 194.65.xxx.xxx/24: medium
○ 194.65.xxx.xxx/24: high
○ 213.13.xxx.xxx/24: high
○ 213.13.xxx.xxx/24: medium
○ 213.13.xxx.xxx/24: low
○ 62.28.xxx.xxx/26: high
```

- **Contexto –** Lista de vulnerabilidades a confrontar com os IPs representadas pelo seu identificador CVE (CVE ID)

```
  - 'CVE-2014-0117'
  - 'CVE-2015-3185'
  - 'CVE-2018-19520'
  - 'CVE-2018-19395'
```

- **Fontes –** Shodan
- **Descrição –** Pesquisa de dispositivos por tipo de vulnerabilidade dada
- **Tarefa –** *rake pull:shodan:case1['altice']*

Após executar a tarefa acima descrita o comportamento da ferramenta foi o seguinte:

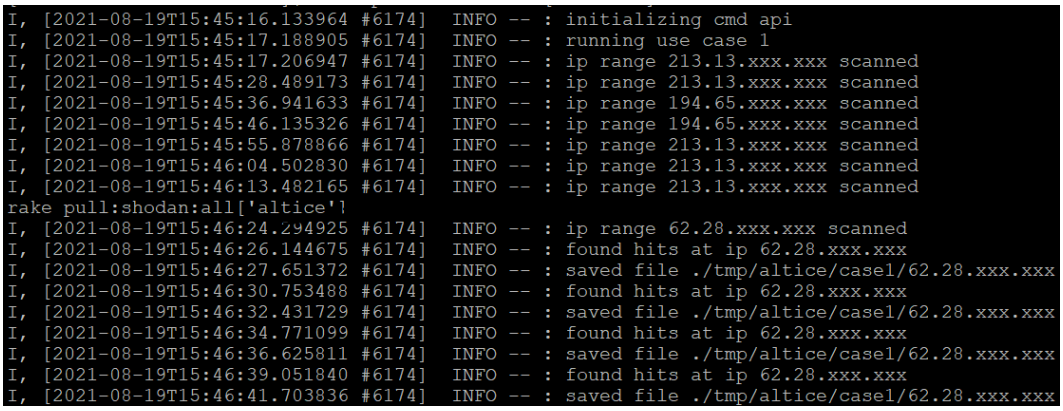


Figura 5.2- Resultado da pesquisa do caso de uso 1

Depois de inicializar a interface de comando do Shodan, como é visível na figura 5.2, uma mensagem de informação é apresentada que dita o início do caso de uso. Após isso, cada intervalo de IPs é analisado como demonstram as mensagens e caso seja encontrada alguma vulnerabilidade da lista é apresenta a mensagem *"found hits at ip x"* e a mensagem de seguida onde confirma que os resultados foram guardados localmente. Com isto comprova-se que este caso de uso funciona em contexto empresarial.

O *dashboard* apresenta uma tabela especifica para este caso de uso como ilustra a figura 5.3 que demonstra que para a primeira entrada, que possui uma versão vulnerável da ferramenta PHP, o seu risco global depois de ser calculado foi seis.

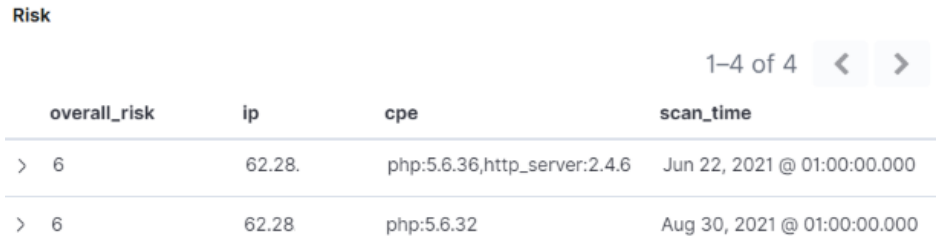


Figura 5.3 - Resultado do dashboard do caso de uso 1

Deteção de dispositivos vulneráveis

Existe uma distinção entre dispositivos com vulnerabilidades e dispositivos vulneráveis. Os primeiros dispositivos são sistemas que podem tornar-se vulneráveis aquando da exploração da vulnerabilidade correspondente e os segundos são sistemas que não têm de ter necessariamente uma vulnerabilidade associada, mas sim falhas de segurança como protocolos mal configurados, portos indesejados abertos ou o uso de palavras-passe de origem dos sistemas. Estes últimos podem ser dispositivos IoT que possam servir como ponto de entrada já que é recorrente existirem falhas de segurança nos protocolos desses produtos. Os *crawlers* do Shodan também ajudam na pesquisa já que caso seja encontrada alguma falha de segurança automaticamente o IP fica associado a uma *tag vulnerable*. Faz sentido, portanto, cobrir ambos os casos fazendo com que este caso de uso também seja fulcral para o contexto da ferramenta.

O caso de uso:

- **Âmbito -** Lista de IPs a pesquisar
- **Contexto –** Lista de palavras chave ou de *banners* específicos de uma má configuração de um protocolo por exemplo

```
  - '220 230 Login successful. port:21'
  - 'tag: vulnerable'
  - 'authentication disabled'
  - 'Port:"445"'
```

- **Fontes –** Shodan e Censys
- **Descrição –** Pesquisa de dispositivos que se encontrem vulneráveis
- **Tarefa -** *rake pull:shodan:case2['altice']*

Após executar a tarefa acima descrita o comportamento da ferramenta foi o seguinte:

```
I, [2021-08-19T15:46:41.704562 #6174]  INFO -- : running use case 2
I, [2021-08-19T15:46:41.706681 #6174]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-19T15:46:46.322536 #6174]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-19T15:46:51.655187 #6174]  INFO -- : ip range 194.65.xxx.xxx scanned
I, [2021-08-19T15:46:57.334622 #6174]  INFO -- : ip range 194.65.xxx.xxx scanned
I, [2021-08-19T15:47:02.797334 #6174]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-19T15:47:07.854520 #6174]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-19T15:47:12.329217 #6174]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-19T15:47:17.474100 #6174]  INFO -- : ip range 62.28.xxx.xxx scanned
```

Figura 5.4 - Resultado da pesquisa do caso de uso 2

Em semelhança ao primeiro caso de uso, na figura 5.2, uma mensagem é apresentada quando o caso de uso é iniciado e a confirmação da pesquisa em cada um dos IPs. Como não existem dispositivos que encaixem na pesquisa efetuada nenhuma mensagem é mostrada. Para

comprovar o funcionamento do caso de uso foi realizada uma pesquisa a um certo intervalo de IPs que são classificados na página *web* do Shodan como vulneráveis foram apresentados resultados, comprovando que o caso de uso funciona e que atualmente a rede exposta da empresa não apresenta dispositivos considerados vulneráveis. Abaixo a figura 5.5 ilustra três IPs que foram considerados vulneráveis já que estão a usar o porto 445 e podem ser infetados mais facilmente já que foi o porto utilizado pelo famoso ataque *WannaCry* [29].

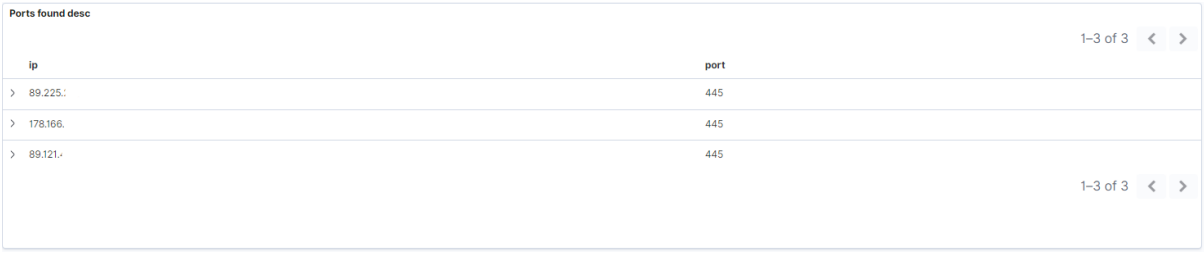


Figura 5.5 - Resultado do dashboard do caso de uso 2

Deteção de Camaras

Para além dos dispositivos considerados vulneráveis, a conexão de camaras IP também pode trazer um ponto de entrada à rede empresarial. Assim, o RoboSCOUT irá pesquisar por dispositivos dado o *banner* apresentado. A existência de vários tipos e marcas de fabricante de camaras também traz desafios na pesquisa das mesmas já que cada uma irá ter um *banner* diferente.

O caso de uso:

- **Âmbito –** Lista de IPs a pesquisar
- **Contexto –** Lista de palavras chave de *banners* de camaras IP

```
  - 'server: hikvision-webs '
  - 'linux upnp avtech'
  - 'Server: SQ-WEBCAM '
  - 'webcamxp'
  - 'yawcam Content-Type: text/html Mime-Type: text/html'
  - 'tag: webcam'
  - 'tag: cam'
  - 'tag: camera'
  - 'IPCamera_Logo'
```

- **Fontes –** Shodan
- **Descrição –** Identificação de camaras na rede
- **Tarefa -** *rake pull:shodan:case3['altice']*

A execução do caso de uso deu o seguinte output:

```
[xfcta25@dev01 RoboSCOUT]$ rake pull:shodan:case3['altice']
I, [2021-08-24T12:48:43.255966 #52032]  INFO -- : initializing cmd api
I, [2021-08-24T12:48:55.413724 #52032]  INFO -- : running use case 3
I, [2021-08-24T12:48:55.421079 #52032]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-24T12:49:23.369191 #52032]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-24T12:49:46.774184 #52032]  INFO -- : ip range 194.65.xxx.xxx scanned
I, [2021-08-24T12:50:31.907661 #52032]  INFO -- : ip range 194.65.xxx.xxx scanned
I, [2021-08-24T12:50:54.439465 #52032]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-24T12:51:16.461523 #52032]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-24T12:51:41.862673 #52032]  INFO -- : ip range 213.13.xxx.xxx scanned
I, [2021-08-24T12:52:00.165059 #52032]  INFO -- : ip range 62.28.xxx.xxx scanned
```

Figura 5.6 - Resultado da pesquisa do caso de uso 3

Também neste caso de uso não é encontrado nenhum resultado que encaixe nos filtros colocados no ficheiro de contexto. Para comprovar o efeito da pesquisa foi realizado o mesmo método que o anterior caso de uso. A figura 5.7 mostra o resultado da pesquisa com o mesmo ficheiro de contexto, mas na rede global o que permite comprovar a funcionalidade da ferramenta para este caso de uso.

Webcams

IP	Location	Port
112.136.	Seoul	23
189.174.	Playa del Carmen	85
61.239.	Hong Kong	3000

Figura 5.7 - Resultado do dashboard do caso de uso 3

Informação Exposta

Na outra vertente do trabalho o uso da ferramenta de RPA permite a pesquisa de informação crítica exposta na Internet. Como é óbvio neste tipo de caso não é possível pesquisar de outra forma como anteriormente para comprovar o efeito dos casos. A única forma é colocar outras pesquisas para comprovar que o processo funciona e capta as informações requisitadas. Para a ferramenta de RPA só é necessário o ficheiro de contexto onde são colocados os termos de pesquisa.

O caso de uso:

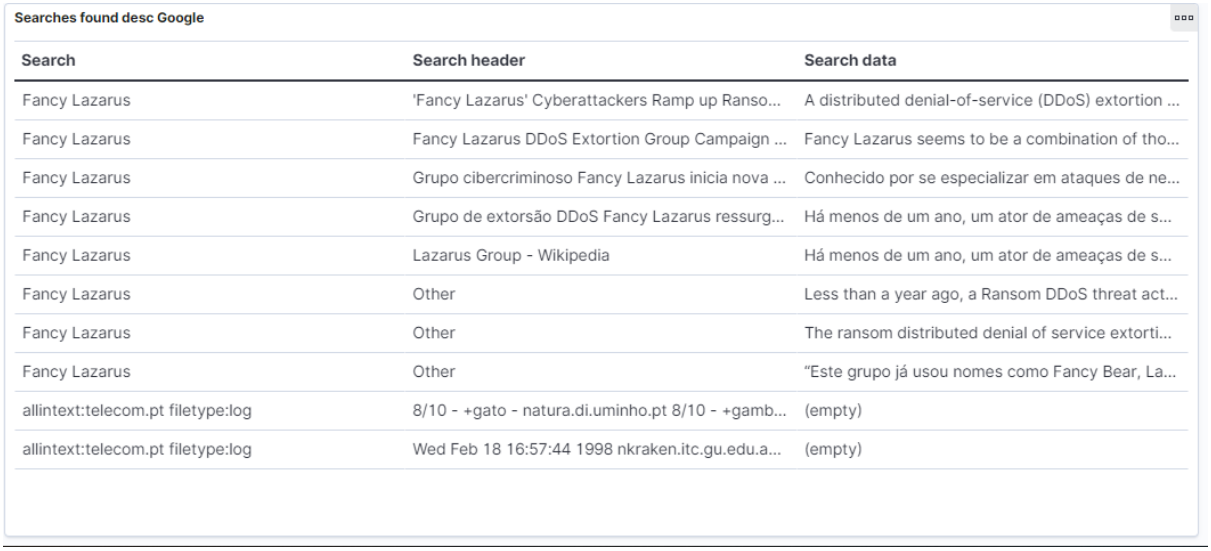
- **Contexto –** Pesquisa a efetuar

```
    Google Dorking:
    "allintext:telecom filetype:pdf",
    "allintext:telecom filetype:txt",
    "altice * pw",
```

```
    "allintext:telecom.pt filetype:log"
    Pastebin:
    "altice",
    "telecom",
    "sapo",
    "telecom.pt",
    "meo"
```

- **Fonte –** Google e Pastebin
- **Descrição –** Pesquisa de informação exposta

Abaixo a figura demonstra um teste efetuado com o Google Dorking e o Pastebin.



Searches found desc Google

Search	Search header	Search data
Fancy Lazarus	'Fancy Lazarus' Cyberattackers Ramp up Ranso...	A distributed denial-of-service (DDoS) extortion ...
Fancy Lazarus	Fancy Lazarus DDoS Extortion Group Campaign ...	Fancy Lazarus seems to be a combination of tho...
Fancy Lazarus	Grupo cibercriminoso Fancy Lazarus inicia nova ...	Conhecido por se especializar em ataques de ne...
Fancy Lazarus	Grupo de extorsão DDoS Fancy Lazarus ressurg...	Há menos de um ano, um ator de ameaças de s...
Fancy Lazarus	Lazarus Group - Wikipedia	Há menos de um ano, um ator de ameaças de s...
Fancy Lazarus	Other	Less than a year ago, a Ransom DDoS threat act...
Fancy Lazarus	Other	The ransom distributed denial of service extorti...
Fancy Lazarus	Other	"Este grupo já usou nomes como Fancy Bear, La...
allintext:telecom.pt filetype:log	8/10 - +gato - natura.di.uminho.pt 8/10 - +gamb...	(empty)
allintext:telecom.pt filetype:log	Wed Feb 18 16:57:44 1998 nkraken.itc.gu.edu.a...	(empty)

Figura 5.8 - Resultado do dashboard do caso de uso 4

Contas Expostas

Para além de informação crítica que possa existir na internet, a existência de contas de colaboradores da empresa, por exemplo, também se torna crítica para o funcionamento da empresa. O objetivo deste caso de uso é dado uma lista de palavras-chave de domínios da empresa, encontrar contas expostas, com ou sem palavra-passe associada. Para encontrar essa informação irão ser utilizadas as mesmas fontes acima referidas.

O caso de uso:

- **Contexto –** Domínios da empresa
- **Fonte –** Google e Pastebin
- **Descrição –** Pesquisa de contas expostas

Neste caso de uso não é possível obter resultados no momento do teste pois não existem correspondências às palavras-chave.

5.2 Conclusão

Com estes cinco casos de uso a ferramenta consegue cobrir um leque de possíveis pontos de entrada no sistema corporativo. Na vertente de sistemas, uma pesquisa periódica ao perímetro da empresa permite com que sejam detetados sistemas vulneráveis, aferir a existência de vulnerabilidades ou a existência de dispositivos externos como camaras. A vertente de informação, por outro lado, permite com que sejam encontrados dados de uma forma atempada que possam comprometer o bom funcionamento empresarial.

Abaixo a Tabela 2 com cada caso de uso, permite analisar o tempo decorrido na busca e processamento da informação bem como os ativos encontrados ou informação que possa ser critica para a empresa.

Caso de Uso	Resultado	Tempo Decorrido	Ativos Encontrados
Deteção de vulnerabilidades	3	27	13
Deteção de dispositivos vulneráveis	3	23	4
Deteção de Camaras	2	11	0
Informação Exposta	3	14	17
Contas Expostas	3	6	0

Tabela 5.1 – Resultado dos casos de uso

Capítulo 6

Conclusão

6.1 Conclusão

Esta tese apresenta um sistema de pesquisa proativa de sistemas vulneráveis e de sistemas vulneráveis e informação crítica exposta na Internet. Este sistema tem como objetivo auxiliar os processos de pesquisa atualmente implementados na empresa.

O sistema foi desenhado de raiz com base nos procedimentos internos da empresa. Foram também criados casos de uso, que serviram como objetivos principais, desde a fase de planeamento para alinhar a arquitetura do RoboSCOUT com os mesmos. Foram criados quatro motores independentes, cada um com o seu objetivo, para permitir a total automatização do sistema só necessitando de ficheiros de entrada que digam aonde pesquisar e o que pesquisar. O motor de calendarização permite coordenar as pesquisas a serem efetuadas para uma melhor adaptação. De forma automática, o motor de pesquisa é inicializado com os ficheiros de âmbito e contexto que permitem a pesquisa ser o mais precisa e concisa possível. A modularização do sistema também permite uma fácil integração de novas fontes sem necessidade de restruturar o *workflow* do mesmo. O motor de processamento efetua uma normalização, agregação dos dados de cada fonte para um único relatório e cálculo do risco de cada ativo que esteja vulnerável, organizando a informação pelas áreas mais importantes. Finalmente, o motor de visualização permite com que os dados em bruto sejam organizados num ambiente mais visualmente apelativo tornando o processo de análise dos mesmos mais eficiente.

A versão inicial do sistema agrega quatro fontes: duas de pesquisa de ativos e duas de pesquisa de informação crítica. Foram realizados testes aos casos de uso inicialmente desenhados e concluiu-se que a eficácia do sistema para os mesmos é garantida.

6.2 Trabalho Futuro

Futuras atualizações ao sistema são essenciais já que as diversas tarefas implementadas podem ser analisadas e melhoradas para permitir uma melhor eficácia do mesmo. A existência de pontos a acrescentar em iterações futuras do sistema são essenciais para o melhoramento dos resultados obtidos das pesquisas efetuadas. Os pontos para essas iterações são:

- **Desenvolvimento de mecanismos de alerta diretos-** Aquando da deteção de informação crítica como contas de utilizador e que possuam palavras-chave associadas ou de ativos que possuam na matriz de risco altas pontuações, deveriam ser comunicadas diretamente, por via email ou WhatsApp, a uma pessoa pré-estabelecida para permitir uma mitigação mais rápida do risco.
- **Integração de uma fonte interna à Altice –** Um sistema que está a ser produzido em paralelo com este é o *0-Day*, que pesquisa por vulnerabilidades de dia 0 e que, aliado ao RoboSCOUT, poderia ser uma fonte para os ficheiros de contexto, analisando a rede para corrigir o problema o mais atempadamente possível.
- **Uso de inteligência artificial no processamento de informação crítica –** Para realmente eliminar falsos positivos da pesquisa dessa informação é necessário a existência de um mecanismo de inteligência artificial que permita uma análise do texto em vez de todos os resultados serem colocados em bruto para serem analisados.

2021

Abreviaturas

IP Internet Protocol.

CSV Comma-separated Values.

CyberSOC Cyber Security Operations Center.

OSINT Open Source Intelligence.

RPA Robotic Process Automation.

API Application Programming Interface.

CLI Command-line Interface.

JSON JavaScript Object Notation.

CVSS Common Vulnerability Scoring System.

CVE Common Vulnerabilities and Exposures.

HTTP Hyper Text Transfer Protocol.

HTTPS Hyper Text Transfer Protocol Secure.

MDAV Motor de Deteção e Análise de Vulnerabilidades.

NVD National Vulnerability Database.

IOT Internet Of Things.

SSH Secure Shell Host.

SFTP SSH File Transfer Protocol.

2021

Referências

[1] “bitsight,” [Online]. Available: https://www.bitsight.com/. [Acedido em 2021].

[2] “qualys,” [Online]. Available: https://www.qualys.com/. [Acedido em 2021].

[3] “cygonito,” [Online]. Available: https://www.cycognito.com/. [Acedido em 2021].

[4] “if you can't beat them, join them,” [Online]. Available: https://en.wiktionary.org/wiki/if_you_can%27t_beat_them,_join_them.

[5] “OSINT,” [Online]. Available: https://pt.wikipedia.org/wiki/OSINT. [Acedido em 16 10 2020].

[6] Censys. [Online]. Available: https://censys.io/overview.

[7] “GitHub-zmap,” [Online]. Available: https://github.com/zmap/zmap.

[8] A. Admininfo.info, «WHAT IT IS, HOW TO USE AND DIFFERENCES BETWEEN ZMAP AND NMAP - TUTORIALS*)*. [Online]. Available: https://en.admininfo.info/qu-es-c-mo-usar-y-diferencias-entre-zmap-y-nmap .

[9] “GitHub-zgrab,” [Online]. Available: https://github.com/zmap/zgrab2.

[10] “ZoomEye,” [Online]. Available: https://www.zoomeye.org/.

[11] “Xmap overview,” [Online]. Available: https://docs.informatica.com/data-integration/b2b-data-transformation/10-5/user-guide/xmap/xmap-overview.html. [Acedido em 12 2020].

[12] “Wmap Overview,” [Online]. Available: https://www.offensive-security.com/metasploit-unleashed/wmap-web-scanner/. [Acedido em 12 2021].

[13] “PasteBin,” [Online]. Available: https://pastebin.com/.

[14] “Ghostbin,” [Online]. Available: https://ghostbin.co/.

[15] “SpiderFoot,” [Online]. Available: https://www.spiderfoot.net/. [Acedido em 14 01 2021].

[16] “Open Data Security,” [Online]. Available: https://opendatasecurity.io/hackers-can-watch-you-via-your-webcam/. [Acedido em 3 11 2020].

[17] P. A. Abdalla e C. Varol, «Testing IoT Security: The Case Study of an IP Camera», em *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, Jun. 2020, pp. 1–5. doi: 10.1109/ISDFS49300.2020.9116392.

[18] C. Bennett, A. Abdou, e P. C. van Oorschot, «Empirical Scanning Analysis of Censys and Shodan», apresentado na Workshop on Measurements, Attacks, and Defenses for the Web, Virtual, 2021. doi: 10.14722/madweb.2021.23009.

[19] “ Common Vulnerabilities and Exposures (CVE),” [Online]. Available: https://cve.mitre.org/.

[20] “Ruby Rakefile,” Ruby Guides, [Online]. Available: https://www.rubyguides.com/2019/02/ruby-rake/. [Acedido em Maio 2021].

[21] “Use Cronjobs,” Opensource, [Online]. Available: https://opensource.com/article/17/11/how-use-cron-linux. [Acedido em 2021].

[22] “Shodan,” [Online]. Available: https://www.shodan.io/. [Acedido em 13 10 2020].

[23] “Elastic Search,” [Online]. Available: https://www.elastic.co/pt/elasticsearch/. [Acedido em 17 11 2020].

[24] “Kibana,” [Online]. Available: https://www.elastic.co/pt/kibana.

[25] “Bundler,” [Online]. Available: https://bundler.io/v2.2/#getting-started. [Acedido em 2021].

[26] “ISO 8601,” [Online]. Available: https://www.iso.org/iso-8601-date-and-time-format.html.

[27] “CVSS v3.1,” [Online]. Available: https://www.first.org/cvss/specification-document.

[28] “NVD,” [Online]. Available: https://nvd.nist.gov/. [Acedido em 2021].

[29] “Avast- Wanna Cry,” [Online]. Available: https://www.avast.com/pt-br/c-wannacry#gref.

[30] “Google Dorks,” [Online]. Available: https://www.welivesecurity.com/br/2021/07/30/google-hacking-verifique-quais-informacoes-sobre-voce-ou-sua-empresa-aparecem-nos-resultados/.

[31] “Banner Grabbing,” [Online]. Available: https://en.wikipedia.org/wiki/Banner_grabbing.

[32] “Camfecting,” [Online]. Available: https://newsroom.unsw.edu.au/news/science-tech/camfecting-how-hackers-attack-gaining-access-your-webcam.

[33] “Shodan Monitor,” [Online]. Available: https://monitor.shodan.io/.

[34] “Robotic Process Automation,” [Online]. Available: https://flow.microsoft.com/en-us/ui-flows/. [Acedido em 2020 11 22].

[35] “BigQuery,” [Online]. Available: https://cloud.google.com/bigquery.

[36] “YAML,” [Online]. Available: https://www.redhat.com/en/topics/automation/what-is-yaml.

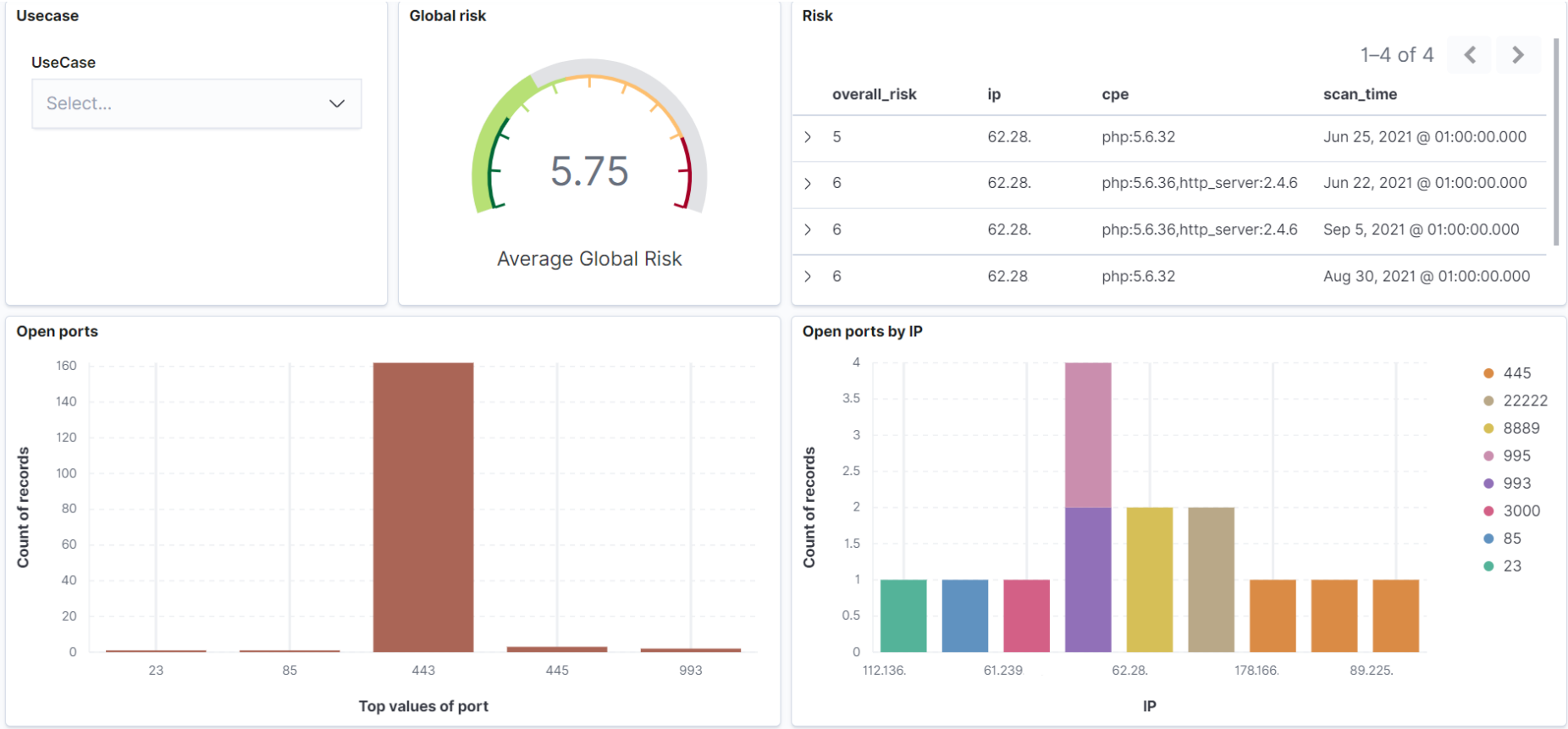
[37] “What is web scraping,” ParseHub, [Online]. Available: https://www.parsehub.com/blog/what-is-web-scraping/.

Apêndice A

Relatório RoboSCOUT

As próximas imagens mostram o painel dos resultados de pesquisa e processamento do sistema. Na primeira parte e segunda é visualizado a componente dos resultados da pesquisa do Shodan e Censys e na terceira a componente da informação coletada por parte do Google e Pastebin.

Exemplo do Relatório – Parte 1



Exemplo do Relatório – Parte 2

IPs without certificate

Top values of ip.keyword	Top values of is_cert_valid
194.65.	false
194.65.	false
194.65.	false
Other	false

Webcams

IP	Location	Port
112.136.	Seoul	23
189.174.	Playa del Carmen	85
61.239.	Hong Kong	3000

Heartbleed

IP	Heartbleed
62.28	7.8

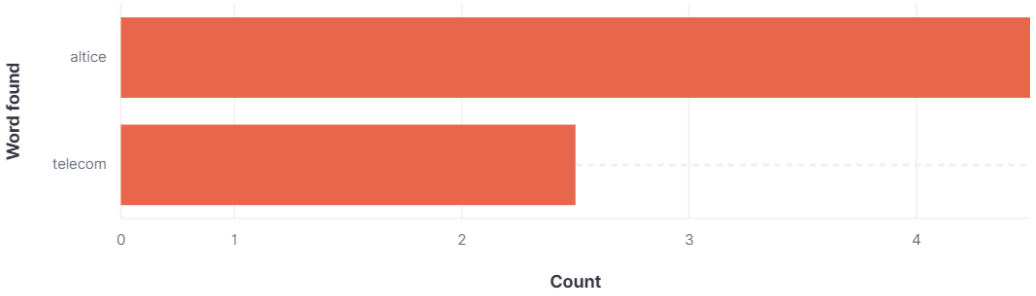
Ports found desc

1–50 of 192

ip	port
62.28.	8889
194.65	443
194.65	443
194.65	443
194.65	443
194.65	443

Exemplo do Relatório – Parte 3

Searches found



Searches found desc PasteBin

Word Hit	Link	Text
altice	https://pastebin.com/MfPwhbie	8.GetEmployeeWithProject
altice	https://pastebin.com/u5GDgJMf	Flower Palace
telecom	https://pastebin.com/Be0JYKAB	Paste Ping

Searches found desc Google

Search	Search header	Search data
Fancy Lazarus	'Fancy Lazarus' Cybe...	A distributed denial-...
Fancy Lazarus	Fancy Lazarus DDoS ...	Fancy Lazarus seem...
Fancy Lazarus	Grupo cibercriminoso...	Conhecido por se es...
Fancy Lazarus	Grupo de extorsão D...	Há menos de um ano...
Fancy Lazarus	Lazarus Group - Wiki...	Há menos de um ano...
Fancy Lazarus	Other	Less than a year ago,...
Fancy Lazarus	Other	The ransom distribut...
Fancy Lazarus	Other	“Este grupo já usou n...
allintext:telecom.pt fil...	8/10 - +gato - natura...	(empty)
allintext:telecom.pt fil...	Wed Feb 18 16:57:44...	(empty)